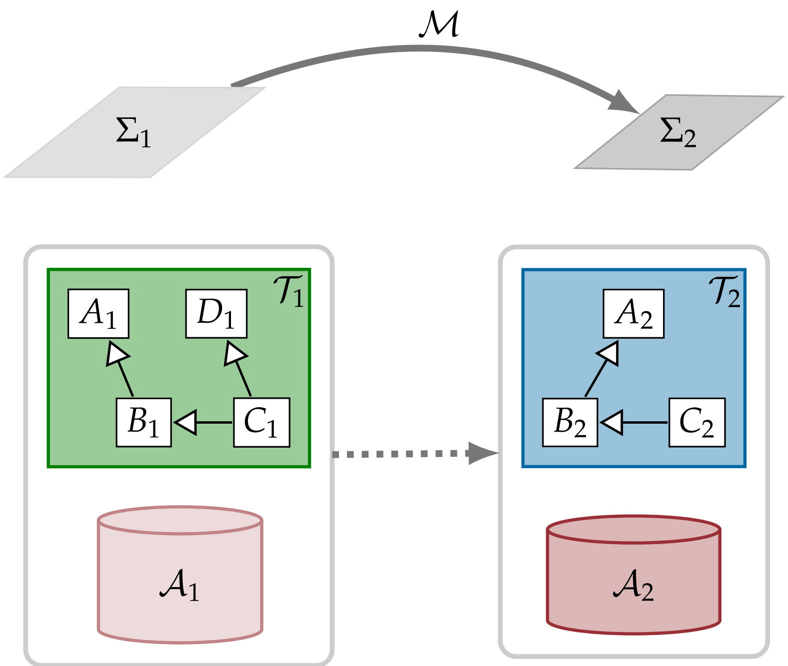
ELENA BOTOEVA

DESCRIPTION LOGIC KNOWLEDGE BASE EXCHANGE



DESCRIPTION LOGIC KNOWLEDGE BASE EXCHANGE

ELENA BOTOEVA

PhD Dissertation

PhilosophiæDoctor (PhD)
Faculty of Computer Science

Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

To my parents Svetlana and Yury,
to my sisters Olga and Maria,
and to my nephews Diana and Sayan.

ABSTRACT

In this thesis, we study the problem of exchanging knowledge between a source and a target knowledge base, connected through mappings. This problem emerges as a fusion of the data exchange problem considered in the traditional database setting on the one hand, and of the knowledge translation problem considered in the knowledge representation and reasoning communities one the other hand: thus, we are interested in exchanging both data and implicit knowledge. As representation formalism we use Description Logics (DLs), thus assuming that the source and target knowledge bases are given as a DL TBox (encoding implicit knowledge) and ABox (encoding data), while the mappings have the form of DL TBox assertions.

To investigate the problem of translating the knowledge in the source knowledge base according to the mapping, we define a general framework of description logic *knowledge base exchange* and specialize it to the case of *DL-Lite*$_{\mathcal{R}}$, a lightweight DL of the *DL-Lite* family. Within this framework we specify three types of translations to be considered that we define as *universal solutions*, *universal* UCQ-*solutions*, and UCQ-*representations*. Universal solutions are the most precise solutions: they preserve all the meaning from the source knowledge base with respect to the mapping. Universal UCQ-solutions is a relaxation of the notion of universal solutions, and they preserve all answers to unions of conjunctive queries (UCQs). UCQ-representations are similar to universal UCQ-solutions, but they do not depend on the source ABox, only on the source TBox and the mapping. The rationale behind the notion of UCQ-representation is to maximize the implicit knowledge in the target, thus, a UCQ-representation of a source TBox captures at best the implicit knowledge that can be extracted from the source according to a mapping using UCQs.

We then develop results for OWL 2 QL, one of the profiles of the standard Web Ontology Language OWL 2, which is based on the DL *DL-Lite*$_{\mathcal{R}}$, and for RDFS, another standard Semantic Web language, which is based on a fragment of *DL-Lite*$_{\mathcal{R}}$ denoted *DL-Lite*$_{RDFS}$. To obtain a good understanding of the knowledge base exchange problem, we study the computational complexity of the *membership* and *non-emptiness* problems for each kind of translation. For universal solutions, e.g., the membership problem checks whether a given candidate target knowledge base is a universal solution for the source knowledge base and the mapping, i.e., whether it belongs to the class of all universal solutions, while the non-emptiness problem answers to the question whether there exists any universal solution for the source knowledge base and the mapping, i.e., whether the class of universal solutions is non-empty. Note that the latter problem is directly related to the task of materializing a translation, moreover, determining UCQ-*representability* is a crucial task, since it allows one to use the obtained target TBox to infer new knowledge in the target, thus reducing the amount of extensional information to be transferred from the source. Adopting a variety of techniques, that include reachability games on graphs and automata on infinite trees, we obtain both upper and lower complexity bounds. For several of the considered cases we are able to precisely characterize the computational complexity of the membership and non-emptiness problems.

PUBLICATIONS

The following publications are related to this PhD work:

Conference Publications

- Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query inseparability for description logic knowledge bases. In *Proc. of the 14th Int. Conf. on Knowledge Representation and Reasoning (KR 2014)*. AAAI Press, 2014
- Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Exchanging OWL 2 QL knowledge bases. In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI 2013)*, pages 703–710, 2013
- Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Exchanging description logic knowledge bases. In *Proc. of the 13th Int. Conf. on Knowledge Representation and Reasoning (KR 2012)*, pages 563–567. AAAI Press, 2012

Workshop Publications

- Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Computing solutions in OWL 2 QL knowledge exchange. In *Proc. of the 26th Int. Workshop on Description Logic (DL 2013)*, volume 1014 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/`, pages 4–16, 2013
- Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Representability in *DL-Lite*$_r$ knowledge base exchange. In *Proc. of the 25th Int. Workshop on Description Logic (DL 2012)*, volume 846 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/`, 2012
- Marcelo Arenas, Elena Botoeva, and Diego Calvanese. Knowledge base exchange. In *Proc. of the 24th Int. Workshop on Description Logic (DL 2011)*, volume 745 of *CEUR Electronic Workshop Proceedings,* `http://ceur-ws.org/`, 2011

Technical Reports

- Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Exchanging OWL 2 QL knowledge bases (extended version). CoRR Technical Report arXiv:1304.5810, arXiv.org e-Print archive, 2013. Available at `http://arxiv.org/abs/1304.5810`

Under Submission

- Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Knowledge base exchange: The case of OWL 2 QL. Under submission to an international journal, 2014

ACKNOWLEDGMENTS

I would like to thank my scientific supervisor Diego Calvanese, my academic father, for everything that he taught me and for being always kind, patient, encouraging and supportive. I am grateful to Marcelo Arenas who has become my second supervisor and without whom the topic of this PhD thesis would have been something different.

I want to thank our Knowledge Representation and Database (KRDB) group and the Free University of Bozen-Bolzano that provided me with all necessary means to carry out my research and, specifically, let me travel to Santiago, Chile where I could start a very fruitful collaboration with Marcelo Arenas.

Also, I would like to express my gratitude to Roman Kontchakov, Michael Zakharyaschev and Frank Wolter for a very enlightening and motivating collaboration. I am very thankful to the reviewers of my PhD thesis, Carsten Lutz and Pablo Barcelò, that provided me with very helpful and constructive comments and that have been and probably will be the only persons who read my entire thesis.

I am sure that accomplishing these PhD studies would not have been possible without my school in my home town Ulan-Ude, Gymnasium 14, my school in Novosibirsk, the Specialized Educational Scientific Center of Novosibirsk State University, and my university in Russia, Novosibirsk State University, where I could get such a great education for free (thanks to my country). In particular, I would like to mention my teachers that contributed to my education: Ul'yana Sergeevna Afanas'eva, Viktor Aleksandrovich Kosteev, Zoya Borisovna Kosteeva, Lyubov' Petrovna Akulova, Larisa Nikolaevna Koreneva, Valeriy Avdeevich Churkin, Aleksandr Ivanovich Kozhanov, Viktor Alekseevich Debelov, Evgeniy Petrov, and many others.

I am very happy that I was surrounded by wonderful people in Bolzano that made and continue making my life here interesting and full of new discoveries, those are Paola Soccio, Yuri Oh, Lucia Veronese and Antonio DeSarro, Magdalena Putz, Oksana Tverdokhlebova, my friends Oksana and Timofey, Babak Bagheri Hariri, Joy and Paul, Naomi and Luis, Tania and Anton, Josef, Evgeny Sherkhonov, Dmitry Milaev, Rumi, Cvetan, Natalia Dikun, Lena Balaeva and Arina, my colleagues at KRDB and at the Faculty of Computer Science.

I am thankful to Providence that brought me to Bolzano and made writing this thesis possible. My endless gratitude and love to my parents for bringing me up the person I am and for being always there for me.

And of course, my special thanks to Vlad for being part of my life.

CONTENTS

Part I

MAIN BODY

1

INTRODUCTION

1.1 MOTIVATION

Ontologies are at the heart of various Computer Science disciplines, among which the most prominent ones are Semantic Web, Biomedical informatics, and of course, Artificial Intelligence and Knowledge Representation. Here, for simplicity, by ontology we mean a formal representation of the knowledge about a domain in terms of axioms expressed over concepts (unary predicates) and roles (binary predicates). In the biomedical domain, e.g., *Pneumonia* and *Lung* could be concepts, and *finding_site* could be a role, and the knowledge about the domain could be asserted in an axiom expressing that *"The finding site of pneumonia is lungs"* [35, 97]. The advantages of using ontologies is that, first, they provide frameworks for organizing and structuring information, and second, it is possible to perform reasoning about the modeled domain.

Usually, some logical language, e.g., first-order logic or predicate logic, is used to formalize ontologies. Each language has an associated expressivity, determined by the logical constructs that can be used; these in turn determine the complexity of reasoning algorithms, i.e., how efficient or inefficient reasoning in that language is. As observed first by Brachman and Levesque [24], there is a tradeoff between expressiveness and complexity of reasoning, i.e., the more expressive the language is, the more inefficient reasoning is. E.g., the standard reasoning task of checking consistency of an ontology written in the horn fragment of predicate logic can be done in polynomial time, while there exists no algorithm at all (i.e., it is *undecidable*) to check whether an ontology written in first-order logic is consistent. Therefore, the task of choosing the ontology language is not trivial, and there has been a lot of research dedicated to studying various languages, such as fragments of first-order logic or extensions of predicate logic, and their computational properties. Nowadays, there exists a whole variety of ontology formalisms. These languages can be roughly divided by their computational complexity into three groups: tractable (first candidates to be used in practice, allow for efficient algorithms), intractable (there exist algorithms, but inefficient), and undecidable (not usable at all). Among the decidable languages, the notable ones are Description Logics, Datalog, and relational databases.

When developing ontologies, one is free to choose the exact terminology and the exact formalism to be used, and there is no single solution to it. For instance, when creating a biomedical ontology about deseases, the lungs can be modeled as *Pair_of_lungs* or *Both_lungs*. Moreover, the ontology developer might choose a description logic, or datalog, or simply a relational database format for modeling and structuring the domain of interest. This led to having complex forms of information, maintained in different formats and organized according to different structures. Often, this information need to be shared between agents: to reuse the existing ontologies, to integrate knowledge from different agents, and so on. Therefore in recent years, both in the data management and in the knowledge representation communities, several settings have been investigated that address this problem from various perspectives:

- in *information integration*, uniform access is provided to a collection of data sources by means of an ontology (or global schema) to which the sources are mapped [83];
- in *peer-to-peer systems*, a set of peers declaratively linked to each other collectively provide access to the information assets they maintain [74, 2, 57];
- in *ontology matching*, the aim is to understand and derive the correspondences between elements in two ontologies [47, 99, 48];
- in *ontology modularity*, the aim is to extract independent, possibly small, subsets of an ontology, so called modules [39, 37, 38];
- in *knowledge translation*, axioms are being translated from one representation (i.e., logical language and vocabulary) into another [28, 42, 43];
- finally, in *data exchange*, the information stored according to a source schema needs to be restructured and translated so as to conform to a target schema [51, 20].

The work we present in this thesis is inspired by the two latter settings, the first one being investigated in knowledge bases and the second one being investigated in databases. In the following, we elaborate on these two settings.

1.1.1 *Knowledge Translation*

The problem of *knowledge translation* has been addressed in 1995 in [28] to formalize the task of reusing/sharing existing encoded knowledge in the process of the development of new intelligent systems, emerged already in the early nineties [56]. An interlingua-based methodology for this problem is proposed, where logical theories encoded in one representation (source) need to be translated to another representation (target) by making use of a first-order logic interlingua. *Interlingua* is a mediating language designed for communicating knowledge between the source and the target representations, and a *representation* is formed using a declarative representation language, a vocabulary and a base theory (associated with the language). Then, the authors devise a formalism for producing translations based on a theory of contexts [88, 27, 64]: a translation is specified as a set of first-order logic sentences each of which describes a rule for deriving a sentence in a target output context that is a translation of a sentence in a source input context. Such an approach, first, provides a formal semantics for translation, and second, enables translation to be done as deduction by a standard theorem prover.

A decade later there has been again an interest in knowledge translation in the context of the Semantic Web, where the problem of communicating knowledge between heterogeneous agents is especially relevant [42, 41, 43]. The focus of these works is to translate axioms represented in a rule-based formalism, where the mapping axioms, that is, the axioms defining how the source and target vocabularies are related, are represented as first-order axioms (although of quite a simple form). The authors devise an

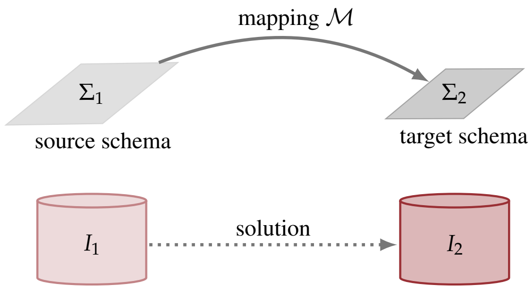


Figure 1: Data Exchange Framework.

algorithm for translating axioms and implement an inference engine that performs the translation.

While the first work [28] is a rather abstract and high level view on the problem of knowledge translation, the other works [42, 41, 43] are more on the practical side and lack solid theoretical foundations. Thus, none of these papers gives a precise understanding of how difficult it is to translate knowledge, for which logical languages it is decidable and for which ones it is undecidable.

1.1.2 *Data Exchange*

Data exchange is a field of database theory, motivated by several applications from industry [98, 65], that deals with transferring data between differently structured databases. The starting point of intensive investigation of the problem of data exchange was given in [51] where it was defined as follows: given data structured according to a source schema and a mapping that provides a declarative account of the relationship between elements of the source schema and elements of a target schema, one wants to transform the source data into data structured according to the target schema so that it accurately reflects the source data with respect to the mapping. The obtained target data instance is referred to as a *solution*. This problem, which can be depicted as in Figure 1, has been extensively studied for different combinations of languages used to specify the source and target schemas, and the mappings [20]. Most of the results in the literature consider source-to-target tuple generating dependencies as the language to specify mappings. *Tuple generating dependencies (tgds)* allow one to express containment between two conjunctive queries: if a conjunction of several predicates holds, then a conjunction of some other predicates must hold, for example,

$$\forall a, b\big(\mathit{AuthorOf}(a,b) \to \exists y, g \,.\, \mathit{BookInfo}(b,a,y) \land \mathit{BookGenre}(b,g)\big), \tag{1}$$

which says that if a is the author of a book b, then there exist y and g such that the information about b is that its author is a and it was written in the year y, and b has genre g. Here, y and g are existentially quantified variables. Many database integrity constraints can be expressed by tgds, so they have been widely employed in all areas of database theory. *Source-to-target tgds (st-tgds)* are tgds of a special shape: the conjunction on the left-hand side uses only symbols from the source schema, while the conjunction on the right-hand side uses only symbols from the target schema.

A fundamental assumption in the traditional data exchange setting is that the source is a complete database: every fact is either true or false. This is not the case for target instances, since incomplete information can be introduced by the mapping layer (see also [84]). As a consequence, for a given source data instance, there can exist many distinct target instances that are solutions for that source data instance.

Example 1.1.1. If we consider the mapping consisting of the constraint in (1), and a source instance consisting of one entry $\mathit{AuthorOf}(\mathsf{tolkien}, \mathsf{lotr})$, encoding that Tolkien is the author of The Lord of the Rings, then the following two target instances are solutions:

$$I_2 = \{\mathit{BookInfo}(\mathsf{lotr}, \mathsf{tolkien}, 1937),\ \mathit{BookGenre}(\mathsf{lotr}, \mathsf{fantasy})\},$$

and

$$I_2' = \{\mathit{BookInfo}(\mathsf{lotr}, \mathsf{tolkien}, \mathrm{NULL}_1),\ \mathit{BookGenre}(\mathsf{lotr}, \mathrm{NULL}_2)\}.$$

Note that here incompleteness is caused by the existential restriction $\exists y, g \ldots$, which can be satisfied by introducing new objects: either named individuals (or constants), like $\mathsf{fantasy}$, or anonymous, like NULL_1. Note also that NULL_1 and NULL_2 are *labeled nulls*, which are the standard way in databases to represent anonymous objects.

To characterize *good* transformations, several criteria have been considered [67]. We emphasize two types of good translations, *universal solutions* and *query solutions*. Universal solutions are the most general solutions: any other solution is more specific (or, detailed). E.g., in Example 1.1.1, I_2' is a universal solution. On the other hand, query solutions are good solutions from the point of view of answering queries formulated over the target schema, so called *target queries*.

Furthermore, the assumption about completeness of the source does not take into account the scenarios in which the source contains some uncertainty. This issue had not been considered until recently, when in [55] the problem of reverse data exchange and in [6] the problem of data exchange with data in the source incompletely specified were addressed. The latter problem is discussed in the next section.

1.1.3 *Data Exchange with Incomplete Information*

In [6], the problem of data exchange with incomplete source data is introduced. Incomplete specification of the source data means that (possibly infinitely) many actual source instances are being represented. A simple example of incomplete data is a database with nulls: consider a table storing information about book genres, and assume that it is known that The Lord of The Rings is a book, but its exact genre is unknown. So this table would contain an entry of the form $\mathit{BookGenre}(\mathsf{lotr}, \mathrm{NULL})$, which represents all different instances containing tuples $\mathit{BookGenre}(\mathsf{lotr}, \mathsf{fantasy})$, $\mathit{BookGenre}(\mathsf{lotr}, \mathsf{history})$, $\mathit{BookGenre}(\mathsf{lotr}, \mathsf{scifi})$, $\ldots$, etc, as depicted in Figure 2. In this setting, when the source is not a complete database, the problem of data exchange becomes substantially more complex. To deal with that, in [6] a general framework for exchanging incomplete information is proposed. This framework is based on the notion of *representation system* as a mechanism to describe in a finite way (infinitely) many complete instances of a data schema.

Knowledge bases are another example of incompletely specified data. A *knowledge base (KB)* is a description of a domain of interest that includes *ground facts*, i.e., information of the form *"John is a student"*, *"Databases is a course"*, *"John attends the Databases course"*, which can be stored in a database (hence extensional information),

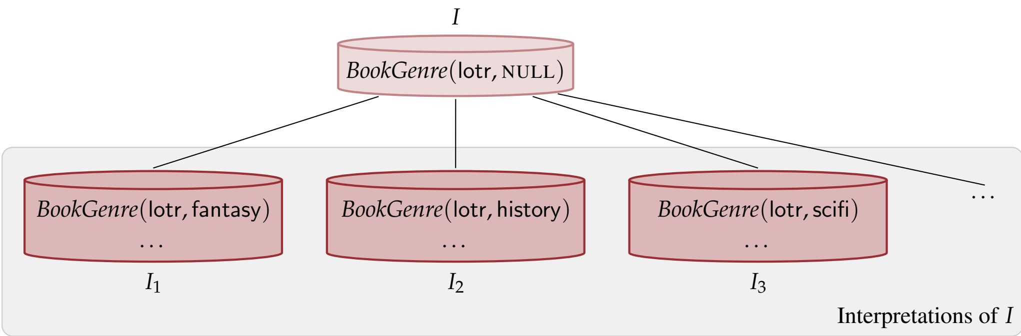


Figure 2: Possible Interpretations of $I = \{BookGenre(\mathsf{lotr}, \textsc{null})\}$.

and *logical axioms*, or what we call an ontology, that structure the knowledge about the domain. E.g., information of the form *"Every course must be taught by somebody"*, *"A student cannot be a professor"* (hence intensional information) . It is intrinsic in the standard semantics of knowledge bases that the knowledge they describe is only a partial description of a domain of interest. It means that a single knowledge base usually represents many actual states of the world. For instance, if we consider the knowledge base consisting of the five axioms mentioned above, then it could represent one possible state of the world, where John also attends the statistics course, David teaches databases and Peter teaches statistics. However, knowledge bases are considerably more expressive than databases with nulls, and even for relatively simple ontology languages, there are knowledge bases that represent infinitely large instances. Importantly, in [6] a general *knowledge exchange framework* is proposed for the case when the source is a knowledge base as opposed to a plain database instance. Then it is shown that already for knowledge bases where the ontological part is expressed by tgds and mappings that are source-to-target tgds, the problem of knowledge base exchange is undecidable, however, if one considers knowledge bases built using *full tgds*, i.e., tgds without existentially quantified variables, then the problem becomes decidable.

To achieve decidability one has to consider less expressive ontological languages. A good candidate for that role is the formalism of Description Logics, which provides fair expressive power, and at the same time possesses good computational properties.

1.1.4 *Description Logics as Ontology Language*

Description Logics (DLs) [17] is a family of formal languages, fragments of first-order logic, specifically designed to serve as ontology languages. They exhibit a reasonable tradeoff between their expressive power and their computational complexity. Nice computational properties in DLs are achieved by restricting attention to unary and binary predicates, called concepts and roles respectively, and to restricted forms of logical axioms. Ground facts in DLs are encoded in the form of an *ABox*, which is a set of membership assertions, and logical axioms are stored in a *TBox*, which is a set of concept and role inclusions. For instance, the DL KB containing the five axioms listed above describing the university domain looks as follows:

$$\begin{array}{ll} Student(\mathsf{john}) & \\ Course(\mathsf{database}) & Course \sqsubseteq \exists teaches^- \\ attends(\mathsf{john}, \mathsf{database}) & Student \sqsubseteq \neg Professor \end{array}$$

where the two inclusions on the right-hand side are concept inclusions, $\textit{teaches}^-$ denotes the inverse of the binary relationship denoted by *teaches*, and $\exists \textit{teaches}^-$ denotes the projection on the second component of *teaches*.

Thus, the main motivation and the starting point for the work done within this thesis are on the one hand, the knowledge translation problem defined in [28], and on the other hand, the knowledge exchange framework defined in [6]. The goal of this PhD work, then, is to investigate the problem of *knowledge base exchange*, where a source KB is connected to a target KB by means of a declarative mapping specification, and the aim is to exchange knowledge from the source to the target by exploiting the mapping. We do so in a setting where the source KB, the target KB, and the mapping are all expressed in variants of Description Logics.

1.2 CONTRIBUTION

This PhD work is concerned with the problem of knowledge base exchange in a Description Logic setting. The contribution of this thesis and the results obtained during my PhD studies can be summarized as follows.

First of all, we propose and develop a framework for KB exchange based on Description Logics (DLs): both source and target are KBs constituted by a DL TBox, representing implicit information, and an ABox, representing explicit information, and mappings are sets of DL concept and role inclusions. We then specialize this framework to the case of lightweight DLs of the *DL-Lite* family [30]. In particular, the most expressive DL we consider is $\textit{DL-Lite}_{\mathcal{R}}$ whose distinguishing feature within the *DL-Lite* family is the presence of role inclusions. In this framework, we are interested in three types of translations that we define as universal solutions, universal UCQ-solutions, and UCQ-representations. *Universal solutions* are the most "precise" solutions: a target KB $\mathcal{K}_2$ is a universal solution for a source KB $\mathcal{K}_1$ under a mapping $\mathcal{M}$ if it preserves all the meaning (i.e., models) of $\mathcal{K}_1$ with respect to $\mathcal{M}$. *Universal* UCQ-*solutions* are a relaxation of the notion of universal solutions: a target KB $\mathcal{K}_2$ is a universal UCQ-solution for a source KB $\mathcal{K}_1$ under a mapping $\mathcal{M}$ if it preserves all answers to unions of conjunctive queries (UCQs) formulated in the target signature and answered over $\mathcal{K}_1$ and $\mathcal{M}$. UCQ-*representations* are similar to universal UCQ-solutions, but they do not depend on the source ABox, only on the source TBox and the mapping: a target TBox $\mathcal{T}_2$ is a UCQ-representation of a source TBox $\mathcal{T}_1$ under a mapping $\mathcal{M}$ if for each possible source ABox $\mathcal{A}_1$, we have that $\mathcal{T}_2$, $\mathcal{M}$ and $\mathcal{A}_1$ give the same answers to UCQs as $\mathcal{T}_1$, $\mathcal{M}$ and $\mathcal{A}_1$. The rationale behind the notion of UCQ-representation is to maximize the implicit knowledge in the target, thus, a UCQ-representation of a source TBox captures at best the intensional information that can be extracted from the source according to a mapping using UCQs.

Secondly, we study and analyze each notion of translation for KBs and mappings defined using the DL $\textit{DL-Lite}_{\mathcal{R}}$. We provide examples that justify the need for target ABoxes with labeled nulls in order for universal solutions and universal UCQ-solutions to exist in some cases, as the language of $\textit{DL-Lite}_{\mathcal{R}}$ is capable of implying existence of new objects. Such ABoxes are called *extended ABoxes* and they can mention anonymous objects implied by the source KB and the mapping by means of labeled nulls, as

Universal solutions	simple ABoxes	extended ABoxes
Membership	PTIME-complete (T 5.2.12)	NP-complete (T 5.3.3)
Non-emptiness	PTIME-complete (T. 5.2.11)	PSPACE-hard (L. 5.3.4)
		in EXPTIME (T. 5.3.9)

Universal UCQ-solutions	simple ABoxes	extended ABoxes
Membership	PSPACE-hard (T. 6.2.1)	PSPACE-hard (T. 6.2.1)
	in EXPTIME (T. 6.2.9)	in EXPTIME (C. 6.3.1)
Non-emptiness		PSPACE-hard (T. 6.3.2)
	in EXPTIME (T. 6.2.10)	

UCQ-**representations**	Complexity
Membership	NLOGSPACE-complete (T. 7.1.8)
Non-emptiness	NLOGSPACE-complete (T. 7.2.12)
Weak UCQ-representability	NLOGSPACE-complete (T. 7.3.5)

Table 1: Complexity results for the membership and non-emptiness problems in $DL\text{-}Lite_{\mathcal{R}}$.

opposed to *simple ABoxes*, which mention exactly the same set of objects as the source ABox. In general, in $DL\text{-}Lite_{\mathcal{R}}$ tree-shaped labeled nulls can be seen as a syntactic sugar, however it is not the case when one cannot freely extend the alphabet or introduce new TBox axioms. Then, we show several cases when universal solutions do not exist, while universal UCQ-solutions do, and that, in general, universal UCQ-solutions display a more robust behavior than universal solutions. For these reasons, we argue that in the context of knowledge base exchange, especially if we consider applications where users only extract information from the translated data by using specific queries (usually conjunctive queries), universal UCQ-solutions and UCQ-representations are the preferred translations over universal solutions.

Finally, we develop results for OWL 2 QL [91], one of the profiles of the standard Web Ontology Language OWL 2 [19], which is based on the DL $DL\text{-}Lite_{\mathcal{R}}$, and for RDFS [25], another standard Semantic Web language, which is based on a fragment of $DL\text{-}Lite_{\mathcal{R}}$ denoted $DL\text{-}Lite_{RDFS}$. To obtain a good understanding of the knowledge base exchange problem, we study the computational complexity of the *membership* and *non-emptiness* problems for each kind of translations and target ABoxes. For universal solutions, e.g., the membership problem checks whether a given candidate target KB is a universal solution for the source KB and the mapping, i.e., whether it belongs to the class of all universal solutions, while the non-emptiness problem answers to the question whether there exists any universal solution for the source KB and the mapping, i.e., whether the class of universal solutions is non-empty. Note that the latter problem is directly related to the task of materializing a translation. Moreover, determining UCQ-*representability* is a crucial task, since it allows one to use the obtained target TBox to infer new knowledge in the target, thus reducing the amount of extensional information

Universal solutions	simple ABoxes	extended ABoxes
Membership	NLOGSPACE-c. (T. 5.4.2)	NP-complete (T. 5.4.3)
Non-emptiness	TRIVIAL (T. 5.4.1)	TRIVIAL (T. 5.4.1)

Universal UCQ-solutions	simple ABoxes	extended ABoxes
Membership	NLOGSPACE-c. (T. 6.4.2)	NP-complete (T. 6.4.3)
Non-emptiness	TRIVIAL (T. 6.4.1)	TRIVIAL (T. 6.4.1)

UCQ-**representations**	Complexity
Membership	NLOGSPACE-complete (T. 7.1.8)
Non-emptiness	NLOGSPACE-complete (T. 7.2.12)
Weak UCQ-representability	TRIVIAL (T. 7.3.6)

Table 2: Complexity results for the membership and non-emptiness problems in $DL\text{-}Lite_{RDFS}$.

to be transferred from the source. For UCQ-representations we also study the problem of *weak* UCQ-*representability*, which answers to the question of whether it is possible to enrich the mapping so as to give a positive answer to the UCQ-representability problem.

The obtained complexity results for $DL\text{-}Lite_{\mathcal{R}}$ can be summarized as in Table 1. For universal solutions with simple ABoxes, we show that both the membership and the non-emptiness problems are decidable in polynomial time by employing techniques based on infinite games on graphs with a reachability acceptance condition (reachability games). We reduce the membership and non-emptiness problems to the problem of finding a winning strategy, which is known to be solvable in PTIME. Then, for universal solutions with extended ABoxes, we prove that the membership problem is NP-complete, while the non-emptiness problem is PSPACE-hard, and provide for the latter an EXPTIME upper bound based on a novel approach exploiting two-way alternating automata. Universal UCQ-solutions show to be more complex, and in contrast to the PTIME upper-bound for the membership problem for universal solutions with simple ABoxes, we show that the membership problem for universal UCQ-solutions with simple ABoxes is already PSPACE-hard. Then, we provide an EXPTIME upper-bound for the membership problem with extended ABoxes by an involved reduction to the reachability games. This algorithm in turn provides us with an EXPTIME algorithm for deciding the non-emptiness problem for universal UCQ-solutions with simple ABoxes. We also show that the non-emptiness problem for universal UCQ-solutions with extended ABoxes is PSPACE-hard. As for UCQ-representations, recall that they do not depend on the shape of the target ABoxes, so we show three complexity bounds for the membership, non-emptiness and weak UCQ-representability problems, which all turn out to be NLOGSPACE-complete.

The complexity results for the case of $DL\text{-}Lite_{RDFS}$ are shown in Table 2. First of all note that in the case of $DL\text{-}Lite_{RDFS}$, the complexity results for universal solutions and universal UCQ-solutions are exactly the same. Importantly, all non-emptiness problems are decidable in constant time; in fact we show that there *always* exist a universal solution (and hence, a universal UCQ-solution), so the answer is trivially "Yes" in each of these cases. Then the membership problems considering simple target ABoxes are NLOGSPACE-complete, and considering extended ABoxes are NP-complete. As for UCQ-representations, the complexity results of the membership and non-emptiness problems are inherited from the case of *DL-Lite*$_{\mathcal{R}}$. Remarkably, the weak UCQ-representability is decidable in constant time, which again in this case means that *DL-Lite*$_{RDFS}$ TBoxes are *always* weakly UCQ-representable.

1.3 STRUCTURE OF THE THESIS

This thesis is structured as follows. We start with presenting in Chapter 2 the preliminary common notions and notations required to read this document. Then, in Chapter 3 we introduce our Knowledge Base exchange framework: we formally define the three notions of translations we are interested in, and set up the space of the complexity problems studied in this thesis. Chapter 4 gives some intuition and basic results about each translation, and provides several illustrative examples when certain translations exist and when they do not. Then, the complexity results and the technical development are presented in Chapter 5 for universal solutions, in Chapter 6 for universal UCQ-solutions, and in Chapter 7 for UCQ-representations. In Chapter 8 we discuss related work, and, in Chapter 9 we summarize and comment on the results obtained in the thesis. We introduce in an appendix some background notions on automata and games that are used in parts of the technical development in the thesis.

2 PRELIMINARIES

In this chapter we introduce the necessary notions and notations employed in the thesis.

2.1 DESCRIPTION LOGICS

Description logics (DLs) [17, 40, 44, 69, 68] is a family of formal languages designed to model a particular domain, and subsequently, to derive new knowledge about it. They provide nice modelling and reasoning capabilities exhibiting a trade-off between expressiveness of a logic and computational complexity of reasoning in that logic.[1] The advances in the field of Description Logics determined their popularity as (underlying logical basis for) formalisms in knowledge representation, ontology-based data access, information and data integration, the Semantic Web, biomedical informatics, etc.

First of all, DLs are languages for building knowledge bases. In DLs, the elements of the domain of interest are structured into *concepts* (unary predicates) and their properties are specified by means of *roles* (binary predicates). General knowledge about the domain of interest is asserted by means of *concept* and *role inclusions* that constitute the *TBox* of a knowledge base ('T' for terminological). Knowledge specific to a particular problem is asserted in the *ABox* ('A' for assertional) in the form of membership assertions, also called facts. Thus, the TBox talks about the intensional level, while the ABox talks about the extensional level. Characteristic features of DLs compared to other modeling formalisms like ER diagrams, UML and datalog, is their concise syntax and intuitive logic-based semantics. The syntax of DLs, which inherits many constructs from the syntax of first-order logic, is variable free and allows to represent complex concept expressions in a compact way. The semantics, which defines the meaning of the information present in the KB, is the standard first-order semantics, where concepts are interpreted as sets of individuals, and roles are interpreted as sets of pairs of individuals.

Second, reasoning is a central service of DLs, it allows one to infer knowledge implicitly represented in the knowledge base. The standard reasoning services over a DL KB include knowledge base satisfiability and query answering. The *knowledge base satisfiability* problem is to check whether the information encoded in the TBox and the ABox is non-contradictory. *Query answering* in knowledge bases is similar to query answering in database only now the information in the TBox should also be taken into account.

Decidability and complexity of reasoning are one of the main concerns in DLs: one is interested in decision procedures (always terminating) that have nice computational properties. Decidability and complexity of the inference problems depend on the expressive power of the DL at hand: in general, the more expressive the logic is, the more complex reasoning becomes. Most of the known DLs are intractable, which means that there exist no efficient (polynomial time) reasoning algorithms. Therefore, there were several lines of research that led to introduction of two families of lightweight DLs with

1 See e.g., `http://www.cs.man.ac.uk/~ezolin/dl/` for a summary of complexity results.

limited expressive power but polynomial algorithms for the basic reasoning tasks: *DL-Lite* [30, 96, 15] and $\mathcal{EL}$ [16, 18]. Most of the result in this thesis are for *DL-Lite*$_{\mathcal{R}}$, a prominent member of the *DL-Lite* family, the logic underlying OWL 2 QL, one of the profiles of OWL 2. In the rest of this section, we define the DLs used later in the thesis and some basic results and notions related to DLs and reasoning with DLs.

2.1.1 *The Description Logic DL-Lite*$_{\mathcal{R}}$ *and Its Sublogics*

The DLs of the *DL-Lite* family [30] of light-weight DLs are characterized by the fact that standard reasoning can be done in polynomial time, and that data complexity of reasoning and conjunctive query answering is in AC^0.

We adapt here *DL-Lite*$_{\mathcal{R}}$, the DL underlying OWL 2 QL, and present now its syntax and semantics. Let N_C, N_R, N_a, N_ℓ be pairwise disjoint sets of *concept names*, *role names*, *constants*, and *labeled nulls*, respectively. Assume in the following that $A \in N_C$ and $P \in N_R$; in *DL-Lite*$_{\mathcal{R}}$, B and C are used to denote basic and arbitrary (or complex) concepts, respectively, and R and Q are used to denote basic and arbitrary (or complex) roles, respectively, defined as follows:

$$\begin{array}{lcl@{\qquad\qquad}lcl} R & ::= & P \mid P^- & B & ::= & A \mid \exists R \\ Q & ::= & R \mid \neg R & C & ::= & B \mid \neg B \end{array}$$

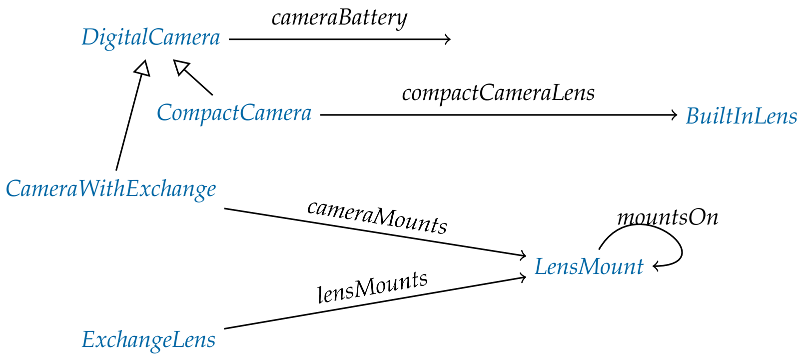
From now on, for a basic role R, we use R^- to denote P^- when $R = P$, and P when $R = P^-$.

A TBox is a finite set of *concept inclusions* $B \sqsubseteq C$ and *role inclusions* $R \sqsubseteq Q$. We call an inclusion of the form $B_1 \sqsubseteq \neg B_2$ or $R_1 \sqsubseteq \neg R_2$ a *disjointness assertion*. An ABox is a finite set of *membership assertions* $B(a)$, $R(a, b)$, where $a, b \in N_a$.

Example 2.1.1. An ontology PhotoCamera about digital photo cameras underlying the structure of an electronics selling website can be described using *DL-Lite*$_{\mathcal{R}}$ syntax.

$$\begin{array}{rcl@{\qquad}l} \mathcal{T}: & & & \mathcal{A}: \\ \mathit{DigitalCamera} & \sqsubseteq & \exists \mathit{cameraBattery} & \mathit{CameraWithExchange}(\mathsf{canon5d}) \\ \mathit{CompactCamera} & \sqsubseteq & \mathit{DigitalCamera} & \mathit{CompactCamera}(\mathsf{lumixFX100}) \\ \mathit{CameraWithExchange} & \sqsubseteq & \mathit{DigitalCamera} & \\ \mathit{CompactCamera} & \sqsubseteq & \neg \mathit{CameraWithExchange} & \\ \mathit{CompactCamera} & \sqsubseteq & \exists \mathit{compactCameraLens} & \\ \exists \mathit{compactCameraLens}^- & \sqsubseteq & \mathit{BuiltInLens} & \\ \mathit{CameraWithExchange} & \sqsubseteq & \exists \mathit{cameraMounts} & \\ \exists \mathit{cameraMounts}^- & \sqsubseteq & \mathit{LensMount} & \\ \mathit{LensMount} & \sqsubseteq & \exists \mathit{lensMounts}^- & \\ \exists \mathit{lensMounts} & \sqsubseteq & \mathit{ExchangeLens} & \end{array}$$

The ER diagram of this TBox can be depicted as follows, where concepts are shown in blue color, arrows of the form $\longrightarrow\!\!\triangleright$ denote "IS A" relation, arrows of the form $\longrightarrow$ with a label R on them denote the binary relationship (i.e., role) R between the corresponding concepts.



■

In this thesis, we also consider extended ABoxes, which are obtained by allowing *labeled nulls* in membership assertions. Observe that in the traditional data exchange setting, labeled nulls are allowed to appear in target instances to be used as placeholders for unknown values [62]. Formally, an *extended ABox* is a finite set of membership assertions $B(u)$ and $R(u,v)$, where $u,v \in (N_a \cup N_\ell)$. Moreover, a(n *extended*) *KB* $\mathcal{K}$ is a pair $\langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a TBox and $\mathcal{A}$ is an (extended) ABox.

In data and knowledge exchange, it is important to distinguish between source and target signatures, and the signatures of different KBs. A *signature* Σ is a finite set of concept and role names, that is, $\Sigma \subseteq N_C \cup N_R$. A KB $\mathcal{K}$ is said to be *defined over* (or simply, *over*) Σ if all the concept and role names occurring in $\mathcal{K}$ belong to Σ (and likewise for TBoxes, ABoxes, concept inclusions, role inclusions and membership assertions). Moreover, the signature of $\mathcal{K}$, denoted by $\Sigma(\mathcal{K})$ is the set of all concept and roles names occurring in $\mathcal{K}$.

DL-Lite$_\mathcal{R}$ is a fragment of first-order logic, so the semantics of *DL-Lite*$_\mathcal{R}$ is inherited from the classical semantics of this logic. Formally, given a signature Σ, an *interpretation* $\mathcal{I}$ *over* Σ is a pair $\langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$, where $\Delta^\mathcal{I}$ is a non-empty domain and $\cdot^\mathcal{I}$ is an interpretation function such that:

(1) $a^\mathcal{I} \in \Delta^\mathcal{I}$, for every constant $a \in N_a$ interpreted by $\mathcal{I}$;

(2) $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$, for every concept name $A \in \Sigma$; and

(3) $P^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, for every role name $P \in \Sigma$.

Function $\cdot^\mathcal{I}$ is extended to also interpret concept and role constructs:

$$
\begin{array}{ll}
(\exists R)^\mathcal{I} = \{x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I} \text{ such that } (x,y) \in R^\mathcal{I}\}; & (\neg B)^\mathcal{I} = \Delta^\mathcal{I} \setminus B^\mathcal{I}; \\
(P^-)^\mathcal{I} = \{(y,x) \in \Delta^\mathcal{I} \times \Delta^\mathcal{I} \mid (x,y) \in P^\mathcal{I}\}; & (\neg R)^\mathcal{I} = (\Delta^\mathcal{I} \times \Delta^\mathcal{I}) \setminus R^\mathcal{I}.
\end{array}
$$

Note that, depending on whether we make the unique name assumption (UNA) or not distinct constants $a,b \in N_a$ may not be allowed to be interpreted as the same object, i.e., $a^\mathcal{I} = b^\mathcal{I}$.

Let $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ be an interpretation over a signature Σ. Then the satisfaction relation $\models$ is defined as follows:

$$
\begin{array}{lll@{\qquad}lll}
\mathcal{I} \models B \sqsubseteq C & \text{if} & B^\mathcal{I} \subseteq C^\mathcal{I} & \mathcal{I} \models B(a) & \text{if} & a^\mathcal{I} \in B^\mathcal{I} \\
\mathcal{I} \models R \sqsubseteq Q & \text{if} & R^\mathcal{I} \subseteq Q^\mathcal{I} & \mathcal{I} \models R(a,b) & \text{if} & (a^\mathcal{I}, b^\mathcal{I}) \in R^\mathcal{I}
\end{array}
$$

$\mathcal{I}$ is said to satisfy a TBox $\mathcal{T}$ over Σ, denoted by $\mathcal{I} \models \mathcal{T}$, if $\mathcal{I} \models \alpha$ for every $\alpha \in \mathcal{T}$. Moreover, satisfaction of membership assertions over Σ is defined as follows. A *substitution* over $\mathcal{I}$ is a partial function $h : (N_a \cup N_\ell) \to \Delta^{\mathcal{I}}$ such that for every $a \in N_a$, $h(a)$ is defined iff $a^{\mathcal{I}}$ is defined, moreover $h(a) = a^{\mathcal{I}}$. Then $\mathcal{I}$ is said to satisfy an (extended) ABox $\mathcal{A}$, denoted by $\mathcal{I} \models \mathcal{A}$, if there exists a substitution h over $\mathcal{I}$ such that:

- for each $B(u) \in \mathcal{A}$, it holds $h(u)$ is defined and $h(u) \in B^{\mathcal{I}}$; and
- for each $R(u,v) \in \mathcal{A}$, it holds $h(u), h(v)$ are defined and $(h(u), h(v)) \in R^{\mathcal{I}}$.

Finally, $\mathcal{I}$ is said to *satisfy* a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, denoted by $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$. Such $\mathcal{I}$ is called a *model* of $\mathcal{K}$, and we use $\textsc{Mod}(\mathcal{K})$ to denote the set of all models of $\mathcal{K}$. We say that $\mathcal{K}$ is *consistent* if $\textsc{Mod}(\mathcal{K}) \neq \emptyset$.

As is customary, given a KB $\mathcal{K}$ over a signature Σ and a membership assertion or an inclusion α over Σ, we use notation $\mathcal{K} \models \alpha$ to indicate that for every interpretation $\mathcal{I}$ of Σ, if $\mathcal{I} \models \mathcal{K}$, then $\mathcal{I} \models \alpha$.

In this thesis we will also consider a fragment of *DL-Lite*$_{\mathcal{R}}$ that corresponds to RDFS [25] and is denoted with *DL-Lite*$_{RDFS}$. Formally, *DL-Lite*$_{RDFS}$ is the fragment of *DL-Lite*$_{\mathcal{R}}$ in which there are only atomic concepts and atomic roles in the right-hand side of inclusions (hence, no disjointness assertions).

2.1.2 *The Canonical Model Property*

The logics of the *DL-Lite* family, and in particular *DL-Lite*$_{\mathcal{R}}$, enjoy the canonical model property. It means that given a KB, if it is consistent, then it is possible to construct a model of that KB with certain characteristics. The first one is that any other model can be obtained from this model. The second one is that it can be used to characterize the certain answers to UCQs (it will be discussed in Chapter 6).

There exist several ways to define the(/a) canonical model. Traditionally in the database theory (in particular, in data exchange), the canonical model of a database instance together with integrity constraints is defined through the notion of *chase* [72, 1], and since a *DL-Lite*$_{\mathcal{R}}$ KB can be seen as such a database, in one of the first papers on *DL-Lite* the canonical model was also defined through the chase [30]. Here we adapt a different definition of the canonical model, which is better suited for the technical development in the scope of this thesis, and can also be found in [77]. Below, we show how the canonical model can be constructed for a *DL-Lite*$_{\mathcal{R}}$ KB. Then, we also define the generating model, which can be used as a small representation of the possibly infinite canonical model.

THE CANONICAL MODEL. Consider a non-extended KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. Denote by $\mathsf{Ind}(\mathcal{A})$ the set of constants occurring in $\mathcal{A}$. Define the equivalence class $[R]$ of a basic role R in $\mathcal{K}$ as

$$[R] = \{S \mid \mathcal{T} \models R \sqsubseteq S \text{ and } \mathcal{T} \models S \sqsubseteq R\}.$$

We introduce a witness $w_{[R]}$ for each equivalence class $[R]$, and write $[R] \leq_{\mathcal{T}} [S]$ if $\mathcal{T} \models R \sqsubseteq S$. Then, the *generating relationship* $\rightsquigarrow_{\mathcal{K}}$ between the set $\mathsf{Ind}(\mathcal{A}) \cup \{w_{[R]} \mid R \text{ is a basic role}\}$ and the set $\{w_{[R]} \mid R \text{ is a basic role}\}$ is defined as follows:

- $a \rightsquigarrow_{\mathcal{K}} w_{[R]}$, if (1) $\mathcal{K} \models \exists R(a)$; (2) $\mathcal{K} \not\models R(a,b)$ for every $b \in \mathsf{Ind}(\mathcal{A})$; (3) $[R'] = [R]$ for every $[R']$ such that $[R'] \leq_{\mathcal{T}} [R]$ and $\mathcal{K} \models \exists R'(a)$.
- $w_{[S]} \rightsquigarrow_{\mathcal{K}} w_{[R]}$, if (1) $\mathcal{T} \models \exists S^- \sqsubseteq \exists R$; (2) $[S^-] \neq [R]$; (3) $[R'] = [R]$ for every $[R']$ such that $[R'] \leq_{\mathcal{T}} [R]$ and $\mathcal{T} \models \exists S^- \sqsubseteq \exists R'$.

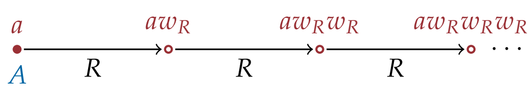
Intuitively, the generating relationship defines when an existing object can be reused to satisfy an axiom of the form $B \sqsubseteq \exists R$, or a new object has to be generated. We denote by $\mathsf{Wit}(\mathcal{K})$ the set of all witnesses $w_{[R]}$ such that for some $a \in \mathsf{Ind}(A)$, $a \rightsquigarrow_{\mathcal{K}} w_{[R_1]} \rightsquigarrow_{\mathcal{K}} \cdots \rightsquigarrow_{\mathcal{K}} w_{[R_n]}$, and $w_{[R_n]} = w_{[R]}$.

Next, we call *$\mathcal{K}$-path* a sequence $aw_{[R_1]} \dots w_{[R_n]}$ such that $n \geq 0$, $a \in N_a$, $a \rightsquigarrow_{\mathcal{K}} w_{[R_1]}$ and $w_{[R_i]} \rightsquigarrow_{\mathcal{K}} w_{[R_{i+1}]}$ for $i \in \{1, \dots, n-1\}$, and denote by $\mathsf{path}(\mathcal{K})$ the set of all $\mathcal{K}$-paths. A $\mathcal{K}$-path $aw_{[R_1]} \dots w_{[R_n]}$ with $n > 0$ encodes one object that has to be generated to satisfy all axioms in $\mathcal{K}$, and is called an *anonymous individual* as it is distinct from any named individual (i.e., constant). Moreover, for every $\sigma \in \mathsf{path}(\mathcal{K})$, denote by $\mathsf{tail}(\sigma)$ the last element in σ. Finally, the *canonical* (or, *universal*) *model* of $\mathcal{K}$, denoted $\mathcal{U}_{\mathcal{K}}$, is defined as:

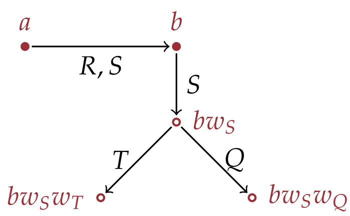
$$
\begin{aligned}
\Delta^{\mathcal{U}_{\mathcal{K}}} &= \mathsf{path}(\mathcal{K}), \\
a^{\mathcal{U}_{\mathcal{K}}} &= a, \text{ for } a \in \mathsf{Ind}(\mathcal{A}), \\
A^{\mathcal{U}_{\mathcal{K}}} &= \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup \{\sigma \cdot w_{[R]} \in \Delta^{\mathcal{U}_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^- \sqsubseteq A\}, \\
P^{\mathcal{U}_{\mathcal{K}}} &= \{(a,b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models P(a,b)\} \cup \\
&\quad\ \{(\sigma, \sigma \cdot w_{[R]}) \mid \mathsf{tail}(\sigma) \rightsquigarrow_{\mathcal{K}} w_{[R]}, [R] \leq_{\mathcal{T}} [P]\} \cup \\
&\quad\ \{(\sigma \cdot w_{[R]}, \sigma) \mid \mathsf{tail}(\sigma) \rightsquigarrow_{\mathcal{K}} w_{[R]}, [R^-] \leq_{\mathcal{T}} [P]\}.
\end{aligned}
$$

Note that the anonymous part of $\mathcal{U}_{\mathcal{K}}$ formed by the anonymous individuals has a tree shape. At the same time, constants can be connected in an arbitrary way. We illustrate graphically several examples of the canonical models below.

Example 2.1.2. Let $\mathcal{K} = \{\mathcal{T}, \mathcal{A}\}$, where $\mathcal{T} = \{A \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists R\}$ and $\mathcal{A} = \{A(a)\}$. Then the canonical model $\mathcal{U}_{\mathcal{K}}$ can be seen as an infinite R-path starting in a, and can depicted according to the following convention, which will be used throughout the thesis: dots and lowercase labels represent domain elements, a label A on a domain element x represents $x \in A^{\mathcal{U}_{\mathcal{K}}}$, and a label R on an arrow between x and y represents that $(x,y) \in R^{\mathcal{U}_{\mathcal{K}}}$:

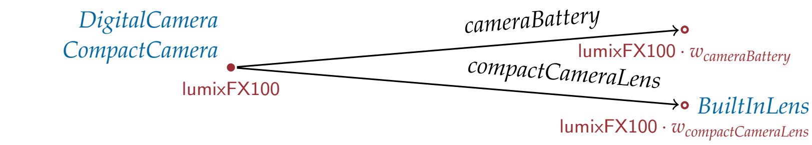


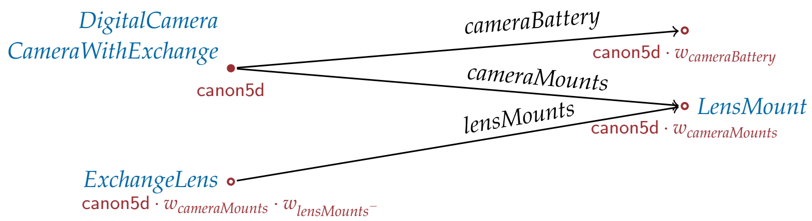
Example 2.1.3. Let $\mathcal{K} = \{\mathcal{T}, \mathcal{A}\}$, where $\mathcal{T} = \{R \sqsubseteq S, \exists R^- \sqsubseteq \exists S, \exists S^- \sqsubseteq \exists T, \exists S^- \sqsubseteq \exists Q\}$ and $\mathcal{A} = \{R(a,b), \exists S(a)\}$. Then the canonical model $\mathcal{U}_{\mathcal{K}}$ can be depicted as follows:



■

Example 2.1.4. Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{T}$ and $\mathcal{A}$ are from Example 2.1.1. Then the canonical model $\mathcal{U}_\mathcal{K}$ can be depicted as below.





■

$\mathcal{U}_\mathcal{K}$ is called the canonical model because every other model of $\mathcal{K}$ is more restricted (or less general) that $\mathcal{U}_\mathcal{K}$. We formalize generality in terms of homomorphisms. For an interpretation $\mathcal{I}$ and a signature Σ, the Σ-*types* $\mathbf{t}^\mathcal{I}_\Sigma(x)$ and $\mathbf{r}^\mathcal{I}_\Sigma(x, y)$ for $x, y \in \Delta^\mathcal{I}$ are given by the set of concepts B (respectively, roles R) over Σ, such that $x \in B^\mathcal{I}$ (respectively, $(x, y) \in R^\mathcal{I}$). We also use $\mathbf{t}^\mathcal{I}(x)$ and $\mathbf{r}^\mathcal{I}(x, y)$ to refer to the types over the signature of all *DL-Lite*$_\mathcal{R}$ concepts and roles.

Given interpretations $\mathcal{I}$ and $\mathcal{J}$, a Σ-*homomorphism* from $\mathcal{I}$ to $\mathcal{J}$ is a function $h : \Delta^\mathcal{I} \mapsto \Delta^\mathcal{J}$ such that

- $h(a^\mathcal{I}) = a^\mathcal{J}$, for all constants a interpreted in $\mathcal{I}$,
- $\mathbf{t}^\mathcal{I}_\Sigma(x) \subseteq \mathbf{t}^\mathcal{J}_\Sigma(h(x))$, and $\mathbf{r}^\mathcal{I}_\Sigma(x, y) \subseteq \mathbf{r}^\mathcal{J}_\Sigma(h(x), h(y))$ for all $x, y \in \Delta^\mathcal{I}$.

We say that $\mathcal{I}$ is Σ-*homomorphically embeddable* into $\mathcal{J}$ if there exists a Σ-homomorphism from $\mathcal{I}$ to $\mathcal{J}$, and $\mathcal{I}$ is Σ-*homomorphically equivalent* to $\mathcal{J}$ if they are Σ-homomorphically embeddable into each other. Intuitively, in the former case, $\mathcal{I}$ is more general than $\mathcal{J}$ with respect to Σ, and in the latter case, $\mathcal{I}$ and $\mathcal{J}$ cannot be distinguished on Σ. If Σ is the set of all *DL-Lite*$_\mathcal{R}$ concepts and roles, we call Σ-homomorphism simply *homomorphism*.

The theorem below establishes the relationship between the canonical model $\mathcal{U}_\mathcal{K}$ and arbitrary models of $\mathcal{K}$.

Theorem 2.1.5 ([77])**.** *If $\mathcal{K}$ is consistent, $\mathcal{U}_\mathcal{K}$ is a model of $\mathcal{K}$. For every model $\mathcal{I}$ of $\mathcal{K}$, there exists a homomorphism from $\mathcal{U}_\mathcal{K}$ to $\mathcal{I}$.*

For an extended KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, denote by $\mathsf{Null}(\mathcal{A})$ the set of labeled nulls occurring in $\mathcal{A}$. Then the canonical model $\mathcal{U}_\mathcal{K}$ is defined by analogy with the construction above, where each labeled null l in $\mathcal{A}$ is interpreted as itself, and the *generating relationship* $\rightsquigarrow_\mathcal{K}$ is defined between the set $\mathsf{Ind}(\mathcal{A}) \cup \mathsf{Null}(\mathcal{A}) \cup \{w_{[R]} \mid R \text{ is a basic role}\}$ and the set $\{w_{[R]} \mid R \text{ is a basic role}\}$. Observe the connection between the labeled nulls in $\mathcal{A}$ and the anonymous individuals in $\Delta^{\mathcal{U}_\mathcal{K}}$: labeled nulls denote exactly such kind of objects, for which it is known they exist, but their exact value (or name, in our case) is not known.

THE GENERATING MODEL. In general, the canonical model of a *DL-Lite*$_{\mathcal{R}}$ KB $\mathcal{K}$ can be infinite, which makes it impossible to deal with it practically. So here we define the generating model of $\mathcal{K}$ that is always finite and can be used for deciding various reasoning tasks efficiently, instead of $\mathcal{U}_{\mathcal{K}}$.

Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an extended KB. Then, the *generating model* of $\mathcal{K}$, denoted $\mathcal{G}_{\mathcal{K}}$, is defined as:

$$
\begin{array}{rcl}
\Delta^{\mathcal{G}_{\mathcal{K}}} & = & \mathsf{Ind}(\mathcal{A}) \cup \mathsf{Null}(\mathcal{A}) \cup \mathsf{Wit}(\mathcal{K}), \\
a^{\mathcal{G}_{\mathcal{K}}} & = & a, \text{ for } a \in \mathsf{Ind}(\mathcal{A}), \\
l^{\mathcal{G}_{\mathcal{K}}} & = & l, \text{ for } l \in \mathsf{Null}(\mathcal{A}), \\
A^{\mathcal{U}_{\mathcal{K}}} & = & \{a \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup \{w_R \in \mathsf{Wit}(\mathcal{K}) \mid \mathcal{T} \models \exists R^- \sqsubseteq A\}, \\
P^{\mathcal{U}_{\mathcal{K}}} & = & \{(a,b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models P(a,b)\} \cup \\
& & \{(x, w_{[R]}) \mid x \leadsto_{\mathcal{K}} w_{[R]}, [R] \leq_{\mathcal{T}} [P]\} \cup \\
& & \{(w_{[R]}, x) \mid x \leadsto_{\mathcal{K}} w_{[R]}, [R^-] \leq_{\mathcal{T}} [P]\}.
\end{array}
$$

It should be clear that $\mathcal{G}_{\mathcal{K}}$ is polynomially large in the size of $\mathcal{K}$. Note that, the canonical model $\mathcal{U}_{\mathcal{K}}$ can be obtained by "unravelling" the generating model $\mathcal{G}_{\mathcal{K}}$.

2.2 QUERIES AND CERTAIN ANSWERS

A k-ary query q over a signature Σ, with $k \geq 0$, denoted $q(x_1, \ldots, x_k)$, is a function that maps every interpretation $\langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ of Σ into a k-ary relation $q^{\mathcal{I}} \subseteq (\Delta^{\mathcal{I}})^k$. In particular, if $k = 0$, then q is said to be a *Boolean query*, denoted $q()$, and $q^{\mathcal{I}}$ is either a relation containing the empty tuple $()$ (representing the value true), then we write $\mathcal{I} \models q$, or the empty relation (representing the value false), in that case we write $\mathcal{I} \not\models q$.

Given a KB $\mathcal{K}$ with $\Sigma \subseteq \Sigma(\mathcal{K})$, the set of *certain answers* to q over $\mathcal{K}$, denoted by $cert(q, \mathcal{K})$, is defined as:

$$
cert(q, \mathcal{K}) = \bigcap_{\mathcal{I} \in \textsc{Mod}(\mathcal{K})} \{(a_1, \ldots, a_k) \mid \{a_1, \ldots, a_k\} \subseteq N_a \text{ and } (a_1^{\mathcal{I}}, \ldots, a_k^{\mathcal{I}}) \in q^{\mathcal{I}}\}.
$$

Each tuple $\vec{a} = (a_1, \ldots, a_k)$ in $cert(q, \mathcal{K})$ is called a *certain answer* for q over $\mathcal{K}$, and we write $\mathcal{K} \models q[\vec{a}]$. Besides, notice that if q is a Boolean query, then $cert(q, \mathcal{K})$ evaluates to true if $q^{\mathcal{I}}$ evaluates to true for every $\mathcal{I} \in \textsc{Mod}(\mathcal{K})$, then we write $\mathcal{K} \models q$, otherwise, it evaluates to false and we write $\mathcal{K} \not\models q$. Notice that the certain answer to a query does *not* contain labeled nulls. Moreover, observe that, if $\mathcal{K}$ is unsatisfiable, then $cert(q, \mathcal{K})$ is trivially the set of all possible tuples $\{(a_1, \ldots, a_k) \mid a_i \in N_a\}$, which we denote by $AllTup(q)$.

The query formalism for which we develop results in this thesis is union of conjunctive queries, a class of well behaving queries widely employed in the database theory and query answering under ontological constraints. First, we define what a conjunctive query and its semantics are. A *conjunctive query* (CQ) *over a signature* Σ is a formula of the form

$$
q(\vec{x}) = \exists \vec{y}.\, \varphi(\vec{x}, \vec{y}),
$$

where $\vec{x}$, $\vec{y}$ are tuples of variables, $\vec{x}$ is the tuple of free variables of $q(\vec{x})$, and $\varphi(\vec{x}, \vec{y})$ is a conjunction of atoms of the form $A(t)$ and $P(t, t')$, where A is a concept name in

Σ, P is a role name in Σ, and each of t, t' is a variable from $\vec{x}$ or $\vec{y}$. Moreover, given an interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ of Σ, the answer of q over $\mathcal{I}$, denoted by $q^{\mathcal{I}}$, is the set of tuples $\vec{a}$ of elements from $\Delta^{\mathcal{I}}$ for which there exist a tuple $\vec{b}$ of elements from $\Delta^{\mathcal{I}}$ such that $\mathcal{I}$ satisfies every conjunct in $\varphi(\vec{a}, \vec{b})$. Finally, a *union of conjunctive queries* (UCQ) over a signature Σ is a first-order formula of the form

$$q(\vec{x}) = \bigvee_{i=1}^{n} q_i(\vec{x}),$$

where each $q_i(\vec{x})$ ($1 \leq i \leq n$) is a CQ over Σ. The semantics of q is defined then as

$$q^{\mathcal{I}} = \bigcup_{i=1}^{n} q_i^{\mathcal{I}}.$$

2.3 COMPLEXITY MEASURES AND COMPLEXITY CLASSES

One of the objectives of this thesis is studying the computational complexity of the reasoning problems that will be formally defined in Section 3.2 and that are *decision problems*, that is, for each instance of a problem the answer should be "Yes" or "No". To assess the complexity we need to specify which *complexity measures* and *complexity classes* we are going to use.

Given a decision problem, the complexity of this problem can be analyzed with respect to different complexity measures depending on which parameters are considered to be the input and which parameters are considered to be fixed. In this thesis we consider only the combined complexity: *combined complexity* is the complexity with respect to the size $|\mathcal{K}|$ of the whole KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, which is equal to $|\mathcal{T}| + |\mathcal{A}|$, that is, the number of symbols used to encode $\mathcal{T}$ and $\mathcal{A}$ respectively.

The complexity of a decision problem is given by the *complexity class* it belongs to. Every complexity class represents a set of problems of related resource-based complexity, where by resource is traditionally understood *time* or *space*, and resources are being used by computation devices, like *Turing machines* [95]. Then, a problem P is said to be *complete* for a complexity class C if P belongs to C and every problem in C can be *reduced* to P. We list the complexity classes relevant for this thesis in the order of increasing complexity:

$$\textsc{Trivial} \subseteq \textsc{NLogSpace} \subseteq \textsc{PTime} \subseteq \textsc{NP} \subseteq \textsc{PSpace} \subseteq \textsc{ExpTime}.$$

For formal definitions of these classes please refer to [58, 81, 22, 104]. Below we give examples of typical complete problems for these classes that will be used to show lower bounds in this these.

NLOGSPACE (non-deterministic logarithmic space) A typical NLOGSPACE-complete problem is *reachability in directed graphs*.

PTIME (polynomial time) A typical PTIME-complete problem is *Circuit Value Problem*.

NP (non-deterministic polynomial time) A typical problem complete for this class is *3-colorability of undirected graphs*.

PSPACE (polynomial space) *Validity of quantified Boolean formula* is a well-known PSPACE-complete problem.

3 KNOWLEDGE BASE EXCHANGE FRAMEWORK

In this chapter we formally define the framework within which this PhD work is done. In Section 3.1, we present an adaptation of the general knowledge exchange framework proposed in [6] to the case of DLs, and in Section 3.2, we define the space of the reasoning problems we interested in solving in this thesis.

3.1 KNOWLEDGE EXCHANGE FRAMEWORK

In this section, we introduce the knowledge exchange framework used in the paper. The starting point to define this framework is the notion of mapping. Assume that Σ_1, Σ_2 are signatures with no concepts or roles in common. An inclusion $E_1 \sqsubseteq E_2$ is said to be *from* Σ_1 *to* Σ_2, if E_1 is a concept or a role over Σ_1 and E_2 is a concept or a role over Σ_2.

Definition 3.1.1. *A* mapping *is a tuple* $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$*, where* $\mathcal{T}_{12}$ *is a TBox consisting of inclusions from* Σ_1 *to* Σ_2.

Intuitively, $\mathcal{M}$ specifies how a KB over the vocabulary Σ_1 should be translated into a KB over the vocabulary Σ_2. Recall that in this thesis, we deal with *DL-Lite*$_\mathcal{R}$ TBoxes, so $\mathcal{T}_{12}$ is assumed to be a set of *DL-Lite*$_\mathcal{R}$ concept and role inclusions.

We continue our digital camera example to illustrate the notion of a mapping.

Example 3.1.2. Assume a different ontology DigitalPhoto talking about digital photo camera which uses the following vocabulary Σ_2:

$$\begin{array}{lll} & \mathit{Battery}(\cdot) & \mathit{requiresBattery}(\cdot,\cdot) \\ \mathit{DigitalPhotoCamera}(\cdot) & \mathit{FixedLens}(\cdot) & \mathit{hasFixedLens}(\cdot,\cdot) \\ \mathit{FixedLensCamera}(\cdot) & \mathit{InterchangeableLens}(\cdot) & \mathit{hasMountType}(\cdot,\cdot) \\ \mathit{InterchangeableLensCamera}(\cdot) & \mathit{MountType}(\cdot) & \mathit{mountsOn}(\cdot,\cdot) \end{array}$$

Then we can specify the relation between the terms in the different ontologies by means of a mapping. More formally, let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where Σ_1 is the vocabulary from Example 2.1.1, and $\mathcal{T}_{12} =$

$$\begin{array}{rlrl} \mathit{DigitalCamera} & \sqsubseteq \mathit{DigitalPhotoCamera} & \mathit{CameraWithExchange} & \\ \mathit{cameraBattery} & \sqsubseteq \mathit{requiresBattery} & & \sqsubseteq \mathit{InterchangeableLensCamera} \\ \exists \mathit{cameraBattery}^- & \sqsubseteq \mathit{Battery} & \mathit{ExchangeLens} & \sqsubseteq \mathit{InterchangeableLens} \\ \mathit{CompactCamera} & \sqsubseteq \mathit{FixedLensCamera} & \mathit{LensMount} & \sqsubseteq \mathit{MountType} \\ \mathit{compactCameraLens} & \sqsubseteq \mathit{hasFixedLens} & \mathit{cameraMounts} & \sqsubseteq \mathit{hasMountType} \\ \mathit{BuiltInLens} & \sqsubseteq \mathit{FixedLens} & \mathit{lensMounts} & \sqsubseteq \mathit{mountsOn} \end{array}$$

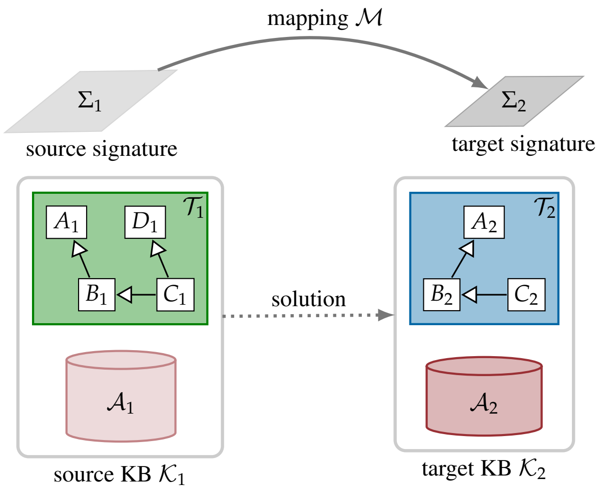


Figure 3: Knowledge Base Exchange Framework.

Thus, $\mathcal{M}$ relates the concepts and roles of PhotoCamera ontology with the concepts and roles of DigitalPhoto ontology. ■

The semantics of a mapping is defined in terms of a notion of satisfaction for interpretations, which has to deal with interpretations not satisfying the UNA (and, more generally, the standard name assumption). More specifically, given interpretations $\mathcal{I}$, $\mathcal{J}$ of Σ_1 and Σ_2, respectively, pair $(\mathcal{I}, \mathcal{J})$ *satisfies* TBox $\mathcal{T}_{12}$, denoted by $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, if

- for every $a \in N_a$ interpreted in $\mathcal{I}$ or $\mathcal{J}$, it holds that $a^{\mathcal{I}} = a^{\mathcal{J}}$,
- for every concept inclusion $B \sqsubseteq C \in \mathcal{T}_{12}$, it holds that $B^{\mathcal{I}} \subseteq C^{\mathcal{J}}$, and
- for every role inclusion $R \sqsubseteq Q \in \mathcal{T}_{12}$, it holds that $R^{\mathcal{I}} \subseteq Q^{\mathcal{J}}$.

Then, $\text{SAT}_{\mathcal{M}}(\mathcal{I})$, the "translation" of $\mathcal{I}$ with respect to $\mathcal{M}$, is defined as the set of interpretations $\mathcal{J}$ of Σ_2 such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, and given a set $\mathcal{X}$ of interpretations of Σ_1, $\text{SAT}_{\mathcal{M}}(\mathcal{X})$ is defined as $\bigcup_{\mathcal{I} \in \mathcal{X}} \text{SAT}_{\mathcal{M}}(\mathcal{I})$.

Notice that the connection between the information in $\mathcal{I}$ and $\mathcal{J}$ is established through the constants that move from source to target according to the mapping. For this reason, we require constants to be interpreted in the same way in $\mathcal{I}$ and $\mathcal{J}$, i.e., to preserve their meaning when they are transferred. This does not hold for labeled nulls, which represent anonymous objects that can be interpreted differently (through different substitutions) in source and target. This distinction between named individuals (i.e., constants) and labeled nulls is important in the context of knowledge exchange.

The main problem studied in the knowledge exchange framework is the problem of translating a KB according to a mapping as it is represented in Figure 3. We formalize this problem through three different notions of translations introduced below (for a thorough comparison of different notions of solutions see Chapter 4).

3.1.1 *Universal Solutions*

The first such notion is the concept of solution and universal solution, which are formalized in the definition below.

Definition 3.1.3. *Given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and KBs $\mathcal{K}_1$, $\mathcal{K}_2$ over Σ_1 and Σ_2, respectively, $\mathcal{K}_2$ is a* solution *for $\mathcal{K}_1$ under $\mathcal{M}$ if*

$$\text{MOD}(\mathcal{K}_2) \subseteq \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1)).$$

Thus, $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$ if every interpretation of $\mathcal{K}_2$ is a valid translation of an interpretation of $\mathcal{K}_1$ according to $\mathcal{M}$. Observe that we require $\mathcal{K}_2$ to be a KB, which implies that the TBox of $\mathcal{K}_2$ is a finite *set of axioms and the ABox of $\mathcal{K}_2$ is a* finite *set of membership assertions.*

Moreover, $\mathcal{K}_2$ is a universal solution *for $\mathcal{K}_1$ under $\mathcal{M}$ if*

$$\text{MOD}(\mathcal{K}_2) = \text{SAT}_{\mathcal{M}}(\text{MOD}(\mathcal{K}_1)).$$

Thus, $\mathcal{K}_2$ is designed to exactly represent the space of interpretations obtained by translating the interpretations of $\mathcal{K}_1$ under $\mathcal{M}$ [7].

We give an example of a universal solution in the scenario of digital cameras.

Example 3.1.4. Let $\mathcal{K}_1 = \langle \mathcal{T}, \mathcal{A}_1 \rangle$ where $\mathcal{T}$ is the TBox of the PhotoCamera KB from Example 2.1.1, $\mathcal{A}_1 = \{CompactCamera(\mathsf{lumixFX100})\}$, and $\mathcal{M}$ the mapping from Example 3.1.2. If we want to talk about Lumix FX 100 in terms of the DigitalPhoto vocabulary and preserve all models, then the following universal solution $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ for $\mathcal{K}_1$ under $\mathcal{M}$, where $\mathcal{T}_2 = \emptyset$ and $\mathcal{A}_2 =$

$FixedLensCamera(\mathsf{lumixFX100})$	$requiresBattery(\mathsf{lumixFX100}, b)$	$Battery(b)$
$DigitalPhotoCamera(\mathsf{lumixFX100})$	$hasFixedLens(\mathsf{lumixFX100}, l)$	$FixedLens(l)$

with fresh labeled nulls b and l, is a candidate KB. ■

For more examples of universal solutions see Section 4.1.

3.1.2 *Universal* UCQ-*Solutions*

A second class of translations is obtained by observing that universal solutions are too restrictive for some applications (see, for instance, Examples 4.1.6 and 4.1.7 for the cases when universal solutions do not exist), in particular when one only needs a translation storing enough information to properly answer some queries. For the particular case of UCQ, this gives rise to the notions of UCQ-solution and universal UCQ-solution.

Definition 3.1.5. *Given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ over Σ_1 and a KB $\mathcal{K}_2$ over Σ_2, $\mathcal{K}_2$ is a* UCQ-solution *for $\mathcal{K}_1$ under $\mathcal{M}$ if for each* UCQ *q over Σ_2:*

$$cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) \subseteq cert(q, \mathcal{K}_2),$$

while $\mathcal{K}_2$ is a universal UCQ-solution *for $\mathcal{K}_1$ under $\mathcal{M}$ if for each* UCQ *q over Σ_2:*

$$cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = cert(q, \mathcal{K}_2).$$

The following is an example of a universal UCQ-solution in the scenario of digital cameras.

Example 3.1.6. Consider $\mathcal{K}_1 = \langle \mathcal{T}, \mathcal{A}_1 \rangle$ from Example 3.1.4. If we want to talk about Lumix FX 100 in terms of the DigitalPhoto vocabulary and preserve only the answers to UCQ, then the following universal UCQ-solution $\mathcal{K}_2' = \langle \mathcal{T}_2', \mathcal{A}_2' \rangle$ for $\mathcal{K}_1$ under $\mathcal{M}$, where $\mathcal{A}_2' = \{FixedLensCamera(\mathsf{lumixFX100})\}$ and $\mathcal{T}_2' =$

$$\begin{array}{ll} DigitalPhotoCamera \sqsubseteq \exists requiresBattery & FixedLensCamera \sqsubseteq DigitalPhotoCamera \\ \exists requiresBattery^- \sqsubseteq Battery & FixedLensCamera \sqsubseteq \exists hasFixedLens \\ & \exists hasFixedLens^- \sqsubseteq FixedLens \end{array}$$

is a candidate KB. ■

Note that the ABoxes and the TBoxes of $\mathcal{K}_2$ and $\mathcal{K}_2'$ from Examples 3.1.4 and 3.1.6 play different roles: $\mathcal{K}_2$ has a big ABox and no TBox, while $\mathcal{K}_2'$ has a small ABox and a big TBox. In Chapter 4 we will go into this difference in more detail.

3.1.3 UCQ-*Representations*

Finally, a last class of translations is obtained by considering that users want to translate as much of the knowledge in a TBox as possible, as a lot of effort is put in practice when constructing a TBox. This observation gives rise to the notion of UCQ-representation [7], which formalizes the idea of translating a source TBox according to a mapping.

Definition 3.1.7. *Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and $\mathcal{T}_1$, $\mathcal{T}_2$ are TBoxes over Σ_1 and Σ_2, respectively. Then $\mathcal{T}_2$ is a* UCQ-representation *of $\mathcal{T}_1$ under $\mathcal{M}$ if for every* UCQ *q over Σ_2 and every ABox $\mathcal{A}_1$ over Σ_1 that is consistent with $\mathcal{T}_1$:*

$$cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) \;=\; cert(q, \langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle). \qquad (\dagger)$$

Notice that $\mathcal{A}_1$ is required to be consistent with $\mathcal{T}_1$ in this definition, to avoid the trivialization of the notion of certain answers because of the use of an inconsistent knowledge base (if $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ is inconsistent, $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = AllTup(q)$, i.e., every possible tuple of constants is in the certain answer).

Below we provide a simple example of a UCQ-representation in the digital camera scenario.

Example 3.1.8. Consider $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ from Example 3.1.2 and a source TBox $\mathcal{T}_1 = \{CompactCamera \sqsubseteq DigitalCamera, CompactCamera \sqsubseteq \exists compactCameraLens\}$. Then TBox $\mathcal{T}_2 = \{FixedLensCamera \sqsubseteq DigitalPhotoCamera, FixedLensCamera \sqsubseteq \exists hasFixedLens\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. ■

To give a better intuition behind the definition of UCQ-representations, assume a graphical presentation of the certain answers to a query q over a KB as the result of applying the TBox arrow to the ABox. Then we can obtain the diagram in Figure 4 illustrating the notion of UCQ-representations. Here the arrows corresponding to $\mathcal{T}_1$ and $\mathcal{T}_2$ are the "standard" isa arrows, and the arrows corresponding to $\mathcal{T}_{12}$ are the wavy arrows. One can see that this diagram is, in fact, a *commutative diagram*, i.e., if $\mathcal{T}_2$ is a UCQ-representation, from the point of view of certain answers, it does not matter if first to follow a $\mathcal{T}_1$-edge and then a $\mathcal{T}_{12}$-edge, or first to follow a $\mathcal{T}_{12}$-edge and then a $\mathcal{T}_2$-edge.

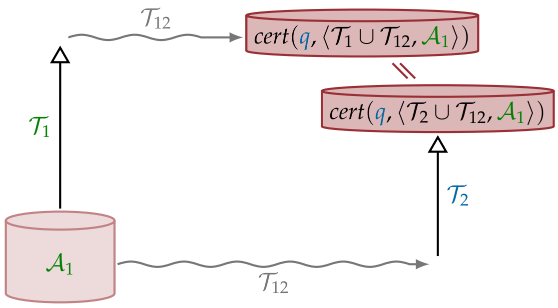


Figure 4: Graphical presentation of a UCQ-representation $\mathcal{T}_2$ of $\mathcal{T}_1$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$.

To conclude this section, we would like to emphasize on why we are interested in UCQ-representations. First of all, UCQ-representations allow to preserve in the target the implicit information from the source, which is conform with the idea of knowledge base exchange as opposed to plain data exchange. Second, UCQ-representations allow to minimize the amount of the extensional information that has to be transferred from the source (which can be large in size), moreover they do not depend on the actual data, so that if the source ABox has been updated it is sufficient to update only the target ABox. Finally, if there exists a UCQ-representation $\mathcal{T}_2$ of a source TBox $\mathcal{T}_1$ under a mapping $\mathcal{M}$, we obtain a straightforward algorithm to construct a universal UCQ-solution for a given source KB $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$: take a target ABox obtained by "translating" the source ABox $\mathcal{A}_1$ with respect to $\mathcal{M}$ and denote it by $\mathcal{M}(\mathcal{A}_1)$, then $\langle \mathcal{T}_2, \mathcal{M}(\mathcal{A}_1) \rangle$ is a universal UCQ-solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$ (see Figure 5). This is how UCQ-representations can "fit" into the knowledge base exchange framework. Observe that $\mathcal{M}(\mathcal{A}_1)$ could be defined as a universal UCQ-solution for $\langle \emptyset, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.

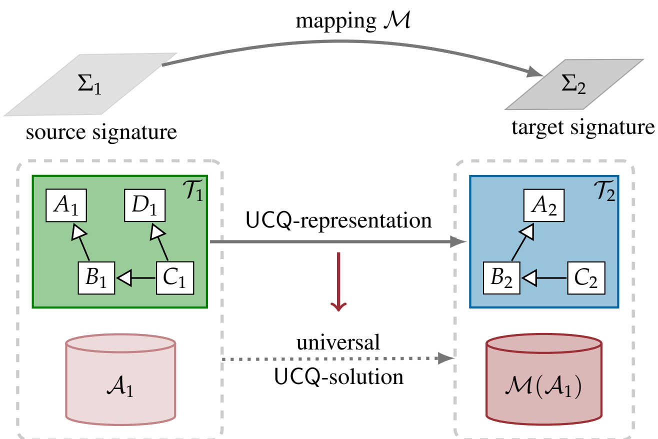


Figure 5: UCQ-representations in the context of Knowledge Base exchange.

3.2 THE SPACE OF REASONING PROBLEMS

In this section we introduce the space of the reasoning problems, whose computational complexity we study in this thesis. Our problem space has three dimensions and can be depicted as in Figure 6.

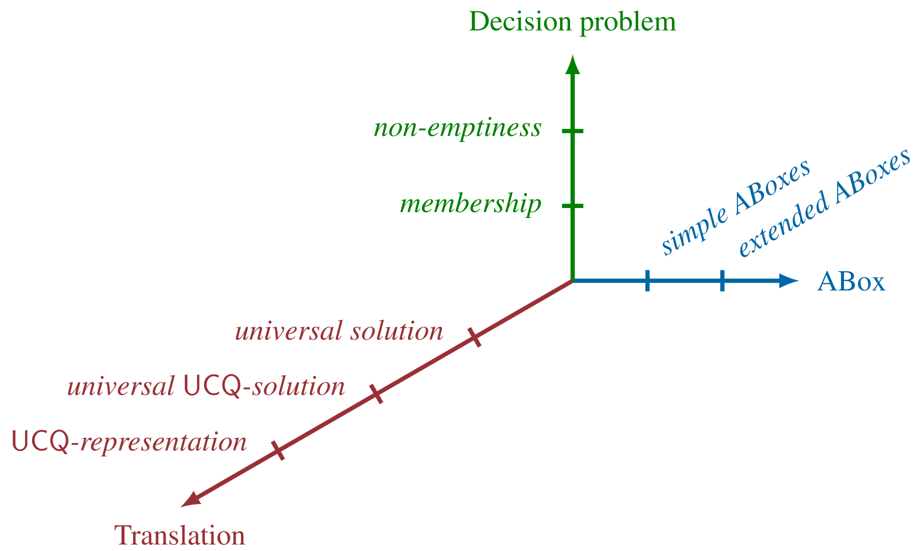


Figure 6: The Space of Reasoning Problems.

First of all, we are interested in the task of computing a translation of a KB or a TBox according to a mapping: it is arguably, the most important problem in knowledge exchange [6, 7], as well as in data exchange [51, 75]. Thus, the first dimension defines the type of translation, and as we presented in the previous section, there are three classes of them: 1) universal solutions, 2) universal UCQ-solutions, and 3) UCQ-representations.

Secondly, as it will become clear in Chapter 4, in order to be able to compute a translation, in some cases it is necessary to use extended ABoxes. Therefore, the second dimension is along the type of ABoxes allowed to be used in translations: 1) simple ABoxes, and 2) extended ABoxes.

Finally, to study the computational complexity of this task for the different notions of translations and target ABoxes, we consider two classical decision problems: the membership problem and the non-emptiness problem, which constitute the third dimension, the decision problem.

1) As usual, the membership problem is concerned with deciding whether a particular instance (a target KB or target TBox, in our case) belongs to the class of the positive instances (all solutions for a given source KB or TBox under a given mapping, in our case). Since we consider three classes of translations: universal solutions, universal UCQ-solutions and UCQ-representations, we are going to deal with three membership problems.

   The *membership* problem for universal solutions (resp. universal UCQ-solutions) has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and KBs $\mathcal{K}_1$, $\mathcal{K}_2$ over Σ_1 and Σ_2, respectively. Then the question to answer is whether $\mathcal{K}_2$ is a universal solution (resp. universal UCQ-solution) for $\mathcal{K}_1$ under $\mathcal{M}$. Moreover, the membership problem for UCQ-representations has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$ over Σ_1 and Σ_2, respectively, and the question to answer is whether $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

2) The non-emptiness problem corresponds to the existential version of the membership problem, and it is concerned with deciding whether there exists any positive instance

(any solution for a given source KB or TBox under a given mapping, in our case). Again, we consider three non-emptiness problems, one for each class of translations.

Formally, the *non-emptiness* problem for universal solutions (resp. universal UCQ-solutions) has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a KB $\mathcal{K}_1$ over Σ_1. Then the question to answer is whether there exists a universal solution (resp. universal UCQ-solution) for $\mathcal{K}_1$ under $\mathcal{M}$. Moreover, the non-emptiness problem for UCQ-representations has as input a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a TBox $\mathcal{T}_1$ over Σ_1, and the question to answer is whether there exists a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. In the case it exists, we say that $\mathcal{T}_1$ is UCQ-*representable* under $\mathcal{M}$, otherwise, $\mathcal{T}_1$ is not UCQ-representable.

In addition to the latter problem, we study the problem of *weak* UCQ-*representability*, which is of interest in the case $\mathcal{T}_1$ is *not* UCQ-representable. So we want to know whether it is possible to "fix" the mapping so that $\mathcal{T}_1$ becomes UCQ-representable. Formally, we say $\mathcal{T}_1$ is *weekly* UCQ-*representable* under $\mathcal{M}$ if there exists a mapping $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ such that $\mathcal{T}_{12} \subseteq \mathcal{T}_{12}^\star$, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^\star$, and $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}^\star$.

Note that the non-emptiness problem is directly related with the problem of computing translations of a KB or a TBox according to a mapping.

Observe that UCQ-representations do not depend on target ABoxes, therefore, in total we study 11 different reasoning problems: 4 for universal solutions, 4 for universal UCQ-solutions and 3 (membership, non-emptiness and weak-representability) for UCQ-representations. In Chapter 5, we investigate the computational complexity of the reasoning problems for universal solutions and present the obtained results. Chapter 6 is dedicated to universal UCQ-solutions and their complexity, and in Chapter 7, we study the notion of UCQ-representability and the reasoning problems associated to it. Note that all results are carried out for *DL-Lite*$_\mathcal{R}$.

4 THE SHAPE OF SOLUTIONS

In this chapter we discuss the notions of translations defined in Section 3.1: each separately in Sections 4.1, 4.2, and 4.3, and then against each other in Section 9.2.

4.1 UNIVERSAL SOLUTIONS

In what follows, we show some known results and examples of universal solutions.

We start with simple examples of universal solutions.

Example 4.1.1. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot)\}, \Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{A \sqsubseteq A',\ B \sqsubseteq B'\}$. Furthermore, assume that $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{\}$ and $\mathcal{A}_1 = \{A(a), B(b)\}$. Then the KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$, where $\mathcal{T}_2 = \emptyset$ and $\mathcal{A}_2 = \{A'(a), B'(b)\}$, is a straightforward universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

Example 4.1.2. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot)\}, \Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{A \sqsubseteq A',\ B \sqsubseteq B'\}$. Furthermore, assume that $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{A \sqsubseteq B\}$ and $\mathcal{A}_1 = \{A(a)\}$. Then the KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$, where $\mathcal{T}_2 = \emptyset$ and $\mathcal{A}_2 = \{A'(a), B'(a)\}$, is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

Universal solutions are the preferred solutions to materialize when exchanging relational databases [51, 52, 20], also in the case of relational databases with incomplete information [6]. However, universal solutions were not thought to take into consideration source data including implicit knowledge (in the form of TBoxes), which is demonstrated in the following example.

Example 4.1.3. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ be as in Example 4.1.2. Furthermore, assume that $\mathcal{K}_2' = \langle \mathcal{T}_2', \mathcal{A}_2' \rangle$, where $\mathcal{T}_2' = \{A' \sqsubseteq B'\}$ and $\mathcal{A}_2' = \{A'(a)\}$. Then we have that $\mathcal{K}_2'$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$. However, we also have that $\mathcal{K}_2'$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. In fact, if $\mathcal{I}$ is an interpretation of Σ_1 such that $\Delta^{\mathcal{I}} = \{a\}$, $A^{\mathcal{I}} = \{a\}$ and $B^{\mathcal{I}} = \{a\}$, and $\mathcal{J}$ is an interpretation of Σ_2 such that $\Delta^{\mathcal{J}} = \{a, b\}$, $B'^{\mathcal{J}} = \{a\}$ and $A'^{\mathcal{J}} = \{a, b\}$, then we have that $\mathcal{I}_1$ is a model of $\mathcal{K}_1$ and $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$ and, therefore, $\mathcal{J} \in \textsc{Sat}_{\mathcal{M}}(\textsc{Mod}(\mathcal{K}_1))$. Thus, we conclude that $\textsc{Sat}_{\mathcal{M}}(\textsc{Mod}(\mathcal{K}_1)) \neq \textsc{Mod}(\mathcal{K}_2')$ as $\mathcal{J}$ is not a model of $\mathcal{K}_2'$ since it does not satisfy inclusion $A' \sqsubseteq B'$. ■

In Examples 4.1.1, 4.1.2 and 4.1.3, a case is shown where universal solutions are not able to represent the implicit source knowledge, as we are only able to construct a universal solution with an empty TBox. In the following proposition, we prove that this is not an isolated phenomenon. In this proposition, we say that a TBox $\mathcal{T}$ over a signature Σ is *trivial* if for every interpretation $\mathcal{I}$ of Σ, it holds that $\mathcal{I} \models \mathcal{T}$ (or, in other words, if $\mathcal{T}$ is equivalent to the empty set of formulas).

Proposition 4.1.4. *Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a DL-Lite$_\mathcal{R}$-mapping, $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ a DL-Lite$_\mathcal{R}$ KB over Σ_1, and $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ a DL-Lite$_\mathcal{R}$ KB over Σ_2. If $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent and $\mathcal{K}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, then $\mathcal{T}_2$ is a trivial TBox.*

Proof. For the sake of contradiction, assume that $\mathcal{T}_2$ is not trivial, that is, there exists an interpretation $\mathcal{J}^\star = \langle \Delta^{\mathcal{J}^\star}, \cdot^{\mathcal{J}^\star} \rangle$ of Σ_2 such that $\mathcal{J}^\star \not\models \mathcal{T}_2$.

Given that $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent, there exists an interpretation $\mathcal{I}^\star = \langle \Delta^{\mathcal{I}^\star}, \cdot^{\mathcal{I}^\star} \rangle$ of $(\Sigma_1 \cup \Sigma_2)$ such that $\mathcal{I}^\star \models \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$. Then define interpretations $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ of Σ_1 and $\mathcal{J} = \langle \Delta^{\mathcal{J}}, \cdot^{\mathcal{J}} \rangle$ of Σ_2 as follows: (1) $\Delta^{\mathcal{I}} = \Delta^{\mathcal{J}} = \Delta^{\mathcal{I}^\star}$; (2) $a^{\mathcal{I}} = a^{\mathcal{J}} = a^{\mathcal{I}^\star}$, for every constant $a \in N_a$; (3) $A_1^{\mathcal{I}} = A_1^{\mathcal{I}^\star}$ and $A_2^{\mathcal{J}} = A_2^{\mathcal{I}^\star}$, for every pair of concept names $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$; and (4) $P_1^{\mathcal{I}} = P_1^{\mathcal{I}^\star}$ and $P_2^{\mathcal{J}} = P_2^{\mathcal{I}^\star}$, for every pair of role names $P_1 \in \Sigma_1$ and $P_2 \in \Sigma_2$. By definition of $\mathcal{I}$, $\mathcal{J}$ and given that $\mathcal{I}^\star \models \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, we conclude that $\mathcal{I} \in \textsc{Mod}(\mathcal{K}_1)$ and $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$.

Without loss generality, we assume that $\Delta^{\mathcal{I}^\star} \cap \Delta^{\mathcal{J}^\star} = \emptyset$. Then define an interpretation $\mathcal{J}'$ of Σ_2 as follows: (1) $\Delta^{\mathcal{J}'} = \Delta^{\mathcal{I}^\star} \cup \Delta^{\mathcal{J}^\star}$; (2) $a^{\mathcal{J}'} = a^{\mathcal{I}^\star}$, for every constant $a \in N_a$; (3) $A^{\mathcal{J}'} = A^{\mathcal{I}^\star} \cup A^{\mathcal{J}^\star}$, for every concept name $A \in \Sigma_2$; and (4) $P^{\mathcal{J}'} = P^{\mathcal{I}^\star} \cup P^{\mathcal{J}^\star}$, for every role name $P \in \Sigma_2$. Given that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, we conclude that $(\mathcal{I}, \mathcal{J}') \models \mathcal{T}_{12}$. In fact, for every concept inclusion $B_1 \sqsubseteq B_2 \in \mathcal{T}_{12}$, where B_1, B_2 are basic concepts, we have that $B_1^{\mathcal{I}} \subseteq B_2^{\mathcal{J}'}$ given that $B_1^{\mathcal{I}} \subseteq B_2^{\mathcal{J}}$, $B_2^{\mathcal{J}} = B_2^{\mathcal{I}^\star}$ and $B_2^{\mathcal{J}'} = B_2^{\mathcal{I}^\star} \cup B_2^{\mathcal{J}^\star}$. Moreover, for every concept inclusion $B_1 \sqsubseteq \neg B_2 \in \mathcal{T}_{12}$, where B_1, B_2 are basic concepts, we have that $B_1^{\mathcal{I}} \subseteq (\neg B_2)^{\mathcal{J}'}$ given that $B_1^{\mathcal{I}} \subseteq (\neg B_2)^{\mathcal{J}}$, $(\neg B_2)^{\mathcal{J}} = (\neg B_2)^{\mathcal{I}^\star}$ and $(\neg B_2)^{\mathcal{J}'} = (\neg B_2)^{\mathcal{I}^\star} \cup (\neg B_2)^{\mathcal{J}^\star}$ (since $B_2^{\mathcal{J}'} = B_2^{\mathcal{I}^\star} \cup B_2^{\mathcal{J}^\star}$ and $\Delta^{\mathcal{I}^\star} \cap \Delta^{\mathcal{J}^\star} = \emptyset$). Finally, for role inclusions $R_1 \sqsubseteq R_2$ and $R_1 \sqsubseteq \neg R_2$ in $\mathcal{T}_{12}$, where R_1, R_2 are basic roles, we conclude that $R_1^{\mathcal{I}} \sqsubseteq R_2^{\mathcal{J}'}$ and $R_1^{\mathcal{I}} \sqsubseteq (\neg R_2)^{\mathcal{J}'}$ as in the previous two cases.

From the results in the previous paragraph, we conclude that $\mathcal{J}' \in \textsc{Sat}_\mathcal{M}(\textsc{Mod}(\mathcal{K}_1))$ (since $\mathcal{J}' \in \textsc{Sat}_\mathcal{M}(\mathcal{I})$ and $\mathcal{I} \in \textsc{Mod}(\mathcal{K}_1)$). On the other hand, we have that $\mathcal{J}' \not\models \mathcal{T}_2$, by definition of $\mathcal{J}'$ and given that $\mathcal{J}^\star \not\models \mathcal{T}_2$. Thus, we have that $\mathcal{J}' \not\models \mathcal{K}_2$ and, thus, $\mathcal{J}' \notin \textsc{Mod}(\mathcal{K}_2)$. Therefore, we conclude that $\textsc{Sat}_\mathcal{M}(\textsc{Mod}(\mathcal{K}_1)) \neq \textsc{Mod}(\mathcal{K}_2)$, which contradicts the fact that $\mathcal{K}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. This concludes the proof of the proposition. □

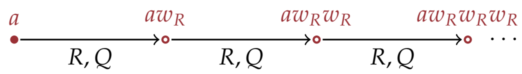
This proposition shows that universal solutions can be viewed as target ABoxes, as the TBox in the generated universal solutions is trivial. Therefore, from now on, in the context of universal solutions, we only consider target KBs of the form $\langle \emptyset, \mathcal{A}_2 \rangle$, and we treat ABoxes $\mathcal{A}_2$ as such KBs.

The following example shows that extended ABoxes are necessary to guarantee the existence of universal solutions in certain cases.

Example 4.1.5. Assume that $\mathcal{M} = \big(\{A(\cdot), R(\cdot,\cdot)\}, \{B(\cdot)\}, \{\exists R^- \sqsubseteq B\}\big)$, and let $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{A \sqsubseteq \exists R\}$ and $\mathcal{A}_1 = \{A(a)\}$. Then a natural way to construct a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$ is to 'populate' the target with all facts implied by $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ (as it is usually done in data exchange). Thus, the ABox $\mathcal{A}_2 = \{B(n)\}$, where n is a labeled null, is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$ if nulls are allowed, which can be readily checked using the definition. Notice that here, a universal solution with non-extended ABoxes does not exist: substituting n by any constant is too restrictive, ruining universality. ■

Our next observation is that there are cases when universal solutions do not exist in OWL 2 QL. This is shown by the following two examples. In the first case, a universal solution does not exist as it is not possible to represent an infinite number of facts in a finite ABox, and the second case illustrates some issues regarding the absence of the UNA, which has to be given up to comply with the OWL 2 QL standard, and regarding the use of disjointness assertions.

Example 4.1.6. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot)\}$, $\Sigma_2 = \{Q(\cdot,\cdot)\}$, and $\mathcal{T}_{12} = \{R \sqsubseteq Q\}$. Furthermore, assume that $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{A}_1 = \{A(a)\}$ and $\mathcal{T}_1 = \{A \sqsubseteq \exists R,\ \exists R^- \sqsubseteq \exists R\}$. In this case, we have that $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is infinite:



so in principle one would need an infinite number of labeled nulls to construct a universal solution. Therefore, if $\mathcal{A}_2$ is an (extended) ABox over Σ_2, then $\mathcal{A}_2$ cannot be a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

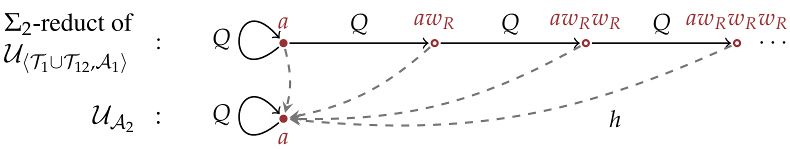
Example 4.1.7. Consider Example 4.1.1 with $\mathcal{T}_1 = \{A \sqsubseteq \neg B\}$. With this seemingly harmless disjointness assertion in $\mathcal{T}_1$, $\mathcal{A}_2$ is no longer a universal solution (not even a solution) for $\mathcal{K}_1$ under $\mathcal{M}$. The reason for that is the lack of the UNA on the one hand, and the presence of the disjointness assertion in $\mathcal{T}_1$ that forces a and b to be interpreted differently in the source, on the other hand. Thus, for a model $\mathcal{J}$ of $\mathcal{A}_2$ such that $a^{\mathcal{J}} = b^{\mathcal{J}}$, $A'^{\mathcal{J}} = B'^{\mathcal{J}} = \{a^{\mathcal{J}}\}$, there is no model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$ (hence, $a^{\mathcal{I}} = a^{\mathcal{J}}$ and $b^{\mathcal{I}} = b^{\mathcal{J}}$). In general, there is no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, even though $\mathcal{K}_1$ and $\mathcal{T}_{12}$ are consistent with each other. ■

We formalize the intuition in the previous examples.

Proposition 4.1.8. *There exists a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ over Σ_1 such that there is no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$.*

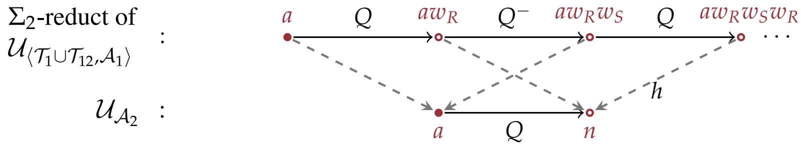
As for the cases when universal solutions do exist, we use the following fact: it can be shown (c.f. Lemma 5.1.2) that in the language without disjointness assertions, an ABox $\mathcal{A}_2$ is a universal solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ if and only if $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Below we illustrate two principal cases when universal solutions exist. In the first case a universal solution exists due to a loop on the ABox constants, and in the second case, due to inverse roles in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$.

Example 4.1.9. Consider Example 4.1.6 with $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot)\}$, $\mathcal{T}_{12} = \{R \sqsubseteq Q, S \sqsubseteq Q\}$ and $\mathcal{A}_1 = \{A(a), S(a,a)\}$. In this case, $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is also infinite, but now there is a loop on a, which allows to deal with all facts implied by $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Consider ABox $\mathcal{A}_2 = \{Q(a,a)\}$. In the following picture, it is easy to see h is a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$. The existence of a Σ_2-homomorphism in the other direction is trivial and, hence, $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$.



■

Example 4.1.10. Consider Example 4.1.6 with $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot)\}$, $\mathcal{T}_{12} = \{R \sqsubseteq Q, S \sqsubseteq Q^-\}$, and $\mathcal{T}_1 = \{A \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists S, \exists S^- \sqsubseteq \exists R\}$. Again, in this example $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is infinite, however, now it is possible to reuse a finite number of nulls to represent all of them, as depicted below. Thus, we have that for a labeled null n, $\mathcal{A}_2 = \{Q(a,n)\}$ is a universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$, if nulls are allowed:



■

Finally, we address the question of how expensive it is to compute universal solutions. We show that universal solutions can be of exponential size for the case of *DL-Lite*$_\mathcal{R}$, thus indicating that it can be difficult to deal with them in practice. In this proposition, $|\mathcal{M}|$ and $|\mathcal{K}|$ are used to denote the sizes of a mapping $\mathcal{M}$ and a KB $\mathcal{K}$, respectively.

Proposition 4.1.11. *There exists a family of mappings* $\{\mathcal{M}_n = (\Sigma_1^n, \Sigma_2^n, \mathcal{T}_{12}^n)\}_{n \geq 1}$ *and a family of KBs* $\{\mathcal{K}_n\}_{n \geq 1}$ *such that every* $\mathcal{K}_n$ *is defined over* Σ_1^n *(*$n \geq 1$*), and the smallest universal solution for* $\mathcal{K}_n$ *under* $\mathcal{M}_n$ *is of size* $2^{\Omega(|\mathcal{M}_n| + |\mathcal{K}_n|)}$.

Proof. Let n be a natural number greater or equal 1. Then $\mathcal{M}^n = (\Sigma_1^n, \Sigma_2^n, \mathcal{T}_{12}^n)$ is defined as follows:

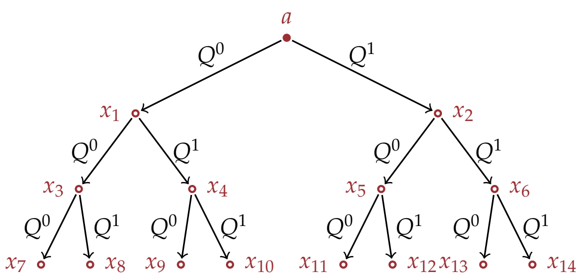
$$\begin{aligned}
\Sigma_1^n &= \{A(\cdot)\} \cup \{R_i^k(\cdot,\cdot) \mid i \in \{1,\dots,n\}, k \in \{0,1\}\},\\
\Sigma_2^n &= \{Q^k(\cdot,\cdot) \mid k \in \{0,1\}\},\\
\mathcal{T}_{12}^n &= \{R_i^k \sqsubseteq Q^k \mid i \in \{1,\dots,n\}, k \in \{0,1\}\},
\end{aligned}$$

and $\mathcal{K}_1^n = \langle \mathcal{T}_1^n, \mathcal{A}_1^n \rangle$, where $\mathcal{T}_1^n$ is the union of the axioms

$$A \sqsubseteq \exists R_1^k, \qquad \exists R_i^{k-} \sqsubseteq \exists R_{i+1}^j$$

for $i \in \{1,\dots,n-1\}$, $k, j \in \{0,1\}$ and $\mathcal{A}_1^n$ consists of one membership assertion $A(a)$.

For every $n \geq 1$, a universal solution $\mathcal{A}_2^n$ for $\mathcal{K}_1^n$ under $\mathcal{M}^n$ exists. This universal solutions $\mathcal{A}_2^n$ is an edge-labeled full binary tree of depth n (that contains 2^n leaves). The root of this tree is a, the label of each edge is one of the role names Q^k, and each node is a labeled null except for the root. Below we depict $\mathcal{A}_2^3$, where $x_1, \dots, x_{14}$ are labeled nulls:



Obviously, $|\mathcal{A}_2^n|$ is of size $2^{\Omega(|\mathcal{M}^n| + |\mathcal{K}_1^n|)}$. We show that $\mathcal{A}_2^n$ is the smallest universal solution for $\mathcal{K}_1^n$ under $\mathcal{M}^n$. For this, we use the fact that universal solutions are homomorphically equivalent to each other (see Lemma 5.1.2) and the following property of graphs:

(CORE) C is a core of a graph A if and only if C is a core, C is a subgraph of A, and there is a homomorphism from A to C,

By contradiction, assume that $\mathcal{A}_2^n$ is not the smallest universal solution, that is, $\mathcal{A}_2^n$ is not a core. Hence there exists a tree $\mathcal{C}$ such that $\mathcal{C}$ is a proper subtree of $\mathcal{A}_2^n$, $\mathcal{C}$ is a core, and there is a homomorphism h from $\mathcal{A}_2^n$ to $\mathcal{C}$. Moreover, h is identity on the elements of $\mathcal{C}$ (Proposition 3.3 in [52]). Let elements $s, t, q \in \mathsf{Ind}(\mathcal{A}_2^n) \cup \mathsf{Null}(\mathcal{A}_2^n)$ be such that s is a parent of t, t is a parent of q (the *parent* relation is defined in the standard way for trees), and $Q^i(s,t) \in \mathcal{A}_2^n$, $Q^j(t,q) \in \mathcal{A}_2^n$ for $i, j \in \{0,1\}$.

First, assume $Q^i(s,t) \in \mathcal{C}$ and $Q^j(t,q) \notin \mathcal{C}$. Therefore, $h(s) = s$, $h(t) = t$ and $h(q)$ must be equal to either to q or to s. In any case we get contradiction with h being a homomorphism from $\mathcal{A}_2^n$ to $\mathcal{C}$: $Q^j \notin \mathbf{r}_{\Sigma_2}^{\mathcal{U}_\mathcal{C}}(t,q) = \{\}$, nor $Q^{j^-} \notin \mathbf{r}_{\Sigma_2}^{\mathcal{U}_\mathcal{C}}(s,t) = \{Q^i\}$.

Secondly, assume $Q^i(s,t) \notin \mathcal{C}$ and $Q^j(t,q) \in \mathcal{C}$. Similarly to the first case, $h(t) = t$ and $h(q) = q$, hence $h(s)$ must be equal to either to s or to q, so h is not a homomorphism.

Finally, assume $Q^0(a, x_1) \in \mathcal{C}$ and $Q^1(a, x_2) \notin \mathcal{C}$. Then, $h(a) = a$, $h(x_1) = x_1$ and $h(x_2)$ must be equal either to x_2 or tox_1, which implies that h is not a homomorphism.

Contradiction rises from the assumption that $\mathcal{A}_2^n$ is not the smallest universal solution. Hence the claim follows. □

4.2 UNIVERSAL UCQ-SOLUTIONS

Our first observation is that the notion of UCQ-solutions is a relaxation of the notion of solutions, hence every universal solution from the previous section is also a universal UCQ-solution. It is formalized in the following proposition.

Proposition 4.2.1. *Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a mapping, $\mathcal{K}_1$ a KB over Σ_1, and $\mathcal{K}_2$ a KB over Σ_2. If $\mathcal{K}_2$ is a (universal) solution for $\mathcal{K}_1$ under $\mathcal{M}$, then $\mathcal{K}_2$ is a (universal)* UCQ*-solution for $\mathcal{K}_1$ under $\mathcal{M}$.*

Proof. Let $\mathcal{K}_2$ be a solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$ and q a UCQ over Σ_2. We show $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) \subseteq cert(q, \mathcal{K}_2)$.

Assume $\mathcal{J}$ is a model of $\mathcal{K}_2$. Since $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$, there exists a model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$. Let $\mathcal{H}$ be the interpretation of $\Sigma_1 \cup \Sigma_2$ defined as the union of $\mathcal{I}$ and $\mathcal{J}$, that is, $\mathcal{H} = \langle \Delta^{\mathcal{H}}, \cdot^{\mathcal{H}} \rangle$, $\Delta^{\mathcal{H}} = \Delta^{\mathcal{I}} \cup \Delta^{\mathcal{J}}$, $a^{\mathcal{H}} = a^{\mathcal{I}}$ for each $a \in N_a$, $A^{\mathcal{H}} = A^{\mathcal{I}}$ for each concept name $A \in \Sigma_1$, $A^{\mathcal{H}} = A^{\mathcal{J}}$ for each concept name $A \in \Sigma_2$, $P^{\mathcal{H}} = P^{\mathcal{I}}$ for each role name $P \in \Sigma_1$, and $P^{\mathcal{H}} = P^{\mathcal{J}}$ for each role name $P \in \Sigma_2$. Then $\mathcal{H}$ is a model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$.

Suppose $\vec{a} \in cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$, it implies $\mathcal{H} \models q(\vec{a})$. Next, as q is a target query, we have that $\mathcal{J} \models q(\vec{a})$. It holds for arbitrary $\mathcal{J}$, therefore we conclude that $\vec{a} \in cert(q, \mathcal{K}_2)$.

Now, let $\mathcal{K}_2$ be a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. In addition, we show $cert(q, \mathcal{K}_2) \subseteq cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$.

Assume $\mathcal{I}$ is a model of $\mathcal{K}_1$ and $\mathcal{J}$ is an interpretation of Σ_2 such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$. Since $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$, it follows $\mathcal{J}$ is a model of $\mathcal{K}_2$. The interpretation $\mathcal{H}$ defined as above is again a model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$.

Suppose $\vec{a} \in cert(q, \mathcal{K}_2)$, it implies $\mathcal{J} \models q(\vec{a})$, and since q is a target query and $\mathcal{J}$ and $\mathcal{H}$ coincide on constants and target symbols, it follows $\mathcal{H} \models q(\vec{a})$. $\mathcal{H}$ is an arbitrary model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, therefore $\vec{a} \in cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$. □

As it is expected, the converse direction does not hold. In our next example we demonstrate a universal UCQ-solution that has a non-trivial TBox, therefore is not a universal solution.

Example 4.2.2. Consider $\mathcal{K}_1$, $\mathcal{M}$ and $\mathcal{K}_2'$ from Example 4.1.3. We have that $\mathcal{K}_2'$ is also a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

Notably, the ABox of $\mathcal{K}_2'$ is smaller than the ABox of the universal solution $\mathcal{K}_2$ (see Example 4.1.2). In many cases, universal UCQ-solutions allow for having smaller ABoxes: there is no need to materialize all facts because they can be derived using axioms in the target TBox.

In Section 4.1 we discussed the cases when universal solutions do not exist. Here we observe that there are cases when universal solutions do not exist but universal UCQ-solutions do. First, an infinite chain can be implied by the target TBox, which is explained in Example 4.2.3. Second, disjointness assertions in the source or the mapping does not have impact on universal UCQ-solutions, which is explained in Example 4.2.4.

Example 4.2.3. Consider $\mathcal{M}$ and $\mathcal{K}_1$ from Example 4.1.6. Recall that there exists no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Instead, KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$, where $\mathcal{T}_2 = \{\exists Q^- \sqsubseteq \exists Q\}$ and $\mathcal{A}_2 = \{\exists Q(a)\}$ is a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. In fact, all and only (up to query equivalence) queries q of the form

$$q = \exists x_1, \cdots, x_{n+1}. Q(x_0, x_1), \cdots, Q(x_n, x_{n+1})$$

where $x_0 = a$, $n \geq 0$ evaluate to true over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$. It is easy to see that the same queries evaluate to true over $\mathcal{K}_2$. ■

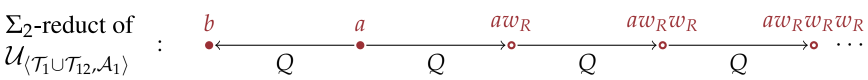
Example 4.2.4. Consider $\mathcal{M}$, $\mathcal{K}_1$, and $\mathcal{A}_2$ from Example 4.1.7. Recall that $\mathcal{A}_2$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. However, notably, $\mathcal{A}_2$ is a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. Moreover, $\mathcal{A}_2$ remains a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$ independently of whether the unique name assumption is employed or not. ■

We formalize the intuition in the previous examples.

Proposition 4.2.5. *There exists a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ over Σ_1 such that there is no universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, while there exists a universal* UCQ*-solution for $\mathcal{K}_1$ under $\mathcal{M}$.*

Unfortunately, there are also cases when universal UCQ-solutions do not exist.

Example 4.2.6. Consider Example 4.1.9 with $\mathcal{A}_1 = \{A(a), S(a, b)\}$. Then the canonical model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ can be depicted as follows



and all and only (up to query equivalence) queries q of the form

$$q = \exists x_1, \cdots, x_{n+1}.Q(x_0, x_1), \cdots, Q(x_n, x_{n+1})$$

and $q = Q(a, b)$, where $x_0 = a$, $n \geq 0$, evaluate to true over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$.

In this case, the basic requirement for a KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ to be a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$ is that $\mathcal{A}_2$ contain $\{\exists Q(a), Q(a, b)\}$. Hence, a similar approach to Example 4.2.3 with having the axiom $\exists Q^- \sqsubseteq \exists Q$ in $\mathcal{T}_2$ does not work, as it would also make the query of the form $\exists x.Q(b, x)$ evaluate to true over $\mathcal{K}_2$, while it evaluates to false over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$. In general, for every TBox $\mathcal{T}_2$ over Σ_2 and every ABox $\mathcal{A}_2$ s.t. $\mathcal{A}_2 \supseteq \{\exists Q(a), Q(a, b)\}$, $\langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is not a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

To continue with the negative results, in some cases we cannot avoid having universal UCQ-solutions of exponential size.

Proposition 4.2.7. *There exists a family of mappings $\{\mathcal{M}_n = (\Sigma_1^n, \Sigma_2^n, \mathcal{T}_{12}^n)\}_{n\geq 1}$ and a family of KBs $\{\mathcal{K}_n\}_{n\geq 1}$ such that every $\mathcal{K}_n$ is defined over Σ_1^n ($n \geq 1$), and the smallest universal* UCQ*-solution for $\mathcal{K}_n$ under $\mathcal{M}_n$ is of size $2^{\Omega(|\mathcal{M}_n|+|\mathcal{K}_n|)}$.*

Proof. Let $n \geq 1$ be a natural number. Then mapping $\mathcal{M}^n = (\Sigma_1^n, \Sigma_2^n, \mathcal{T}_{12}^n)$ is defined as follows:

$$\begin{array}{rcl} \Sigma_1^n & = & \{A(\cdot)\} \cup \{R_i^k(\cdot,\cdot), S^k(\cdot,\cdot) \mid i \in \{1, \ldots, n\}, k \in \{0, 1\}\}, \\ \Sigma_2^n & = & \{Q^k(\cdot,\cdot) \mid k \in \{0, 1\}\}, \\ \mathcal{T}_{12}^n & = & \{R_i^k \sqsubseteq Q^k, S^k \sqsubseteq Q^k \mid i \in \{1, \ldots, n\}, k \in \{0, 1\}\}. \end{array}$$

Moreover, knowledge base $\mathcal{K}_1^n$ is defined as $\langle \mathcal{T}_1^n, \mathcal{A}_1^n \rangle$, where $\mathcal{T}_1^n$ is defined as the union of the axioms

$$A \sqsubseteq \exists R_1^k, \qquad \exists R_i^{k-} \sqsubseteq \exists R_{i+1}^j$$

for $i \in \{1, \ldots, n-1\}$, $k, j \in \{0, 1\}$, and $\mathcal{A}_1^n$ is defined as $\{A(a), S^0(b, c), S^1(d, e)\}$, where a, b, c, d, e are pairwise distinct constants.

For every $n \geq 1$, a universal solution $\mathcal{A}_2^n$ for $\mathcal{K}_1^n$ under $\mathcal{M}^n$ exists. This universal solution $\mathcal{A}_2^n$ consists of membership assertions $Q^0(b, c), Q^1(d, e)$ together with an edge-labeled full binary tree of depth n (that contains 2^n leaves). As in the proof of Proposition 4.1.11, the root of this tree is a, the label of each edge is one of the role names Q^k ($k \in \{0, 1\}$), and the tree contains labeled nulls in every node except for the root.

In this case, there exist no universal UCQ-solution distinct from the universal solutions for $\mathcal{K}_1^n$ under $\mathcal{M}^n$, as none of the non-trivial axioms over $\Sigma_2^n = \{Q^0, Q^1\}$ can be present in the target. Hence, as in the proof of Proposition 4.1.11, we can conclude that $\mathcal{A}_2^n$ is the smallest universal UCQ-solution for $\mathcal{K}_1^n$ under $\mathcal{M}^n$, from which the proposition follows. □

4.3 UCQ-REPRESENTATIONS

We start by discussing several examples where it becomes clear how UCQ-representations depend on the mapping. First, we consider signatures with concepts only.

Example 4.3.1. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot), C(\cdot)\}$ and $\Sigma_2 = \{A'(\cdot), B'(\cdot), C'(\cdot)\}$. Moreover, let $\mathcal{T}_1 = \{A \sqsubseteq B\}$. Consider the following cases of $\mathcal{T}_{12}$:

(1) $\mathcal{T}_{12} = \{B \sqsubseteq B'\}$.
Then there exists no UCQ-representation: take ABox $\mathcal{A}_1 = \{A(a)\}$, then query $q = B'(a)$ evaluates to true over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$. However, for every target TBox $\mathcal{T}_2$, q evaluates to false over $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$.

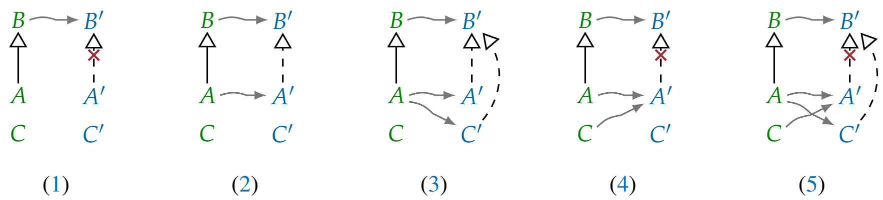
(2) $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B'\}$.
Then as expected, $\mathcal{T}_2 = \{A' \sqsubseteq B'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

(3) $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', A \sqsubseteq C'\}$.
Then, there exist several UCQ-representations: $\mathcal{T}_2 = \{A' \sqsubseteq B'\}$, $\mathcal{T}_2' = \{C' \sqsubseteq B'\}$ and their combination.

(4) $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq A'\}$.
Then, there exists no UCQ-representation: on the one hand, if a target TBox contains $A' \sqsubseteq B'$, then for $\mathcal{A}_1 = \{C(c)\}$, $q = B'(c)$ evaluates to true over $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$ and to false over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$. On the other hand, if a target TBox does not contain $A' \sqsubseteq B'$, then for $\mathcal{A}_1 = \{A(a)\}$, $q = B'(a)$ evaluates to true over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$ and to false over $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$.

(5) $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', A \sqsubseteq C', C \sqsubseteq A'\}$.
Then, $\mathcal{T}_2' = \{C' \sqsubseteq B'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. Note that $\mathcal{T}_2 = \{A' \sqsubseteq B'\}$ is not a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ for the same reason as explained in item (4).

These cases can be depicted in the following ER diagrams, where source concepts are shown in green, target concepts in blue, and the mapping connections are drawn with "wavy" arrows:



∎

In the next example we also consider roles.

Example 4.3.2. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot)\}$ and $\Sigma_2 = \{A'(\cdot), R'(\cdot,\cdot), B'(\cdot)\}$. Moreover, let $\mathcal{T}_1 = \{A \sqsubseteq \exists R\}$. Consider the following cases of $\mathcal{T}_{12}$:

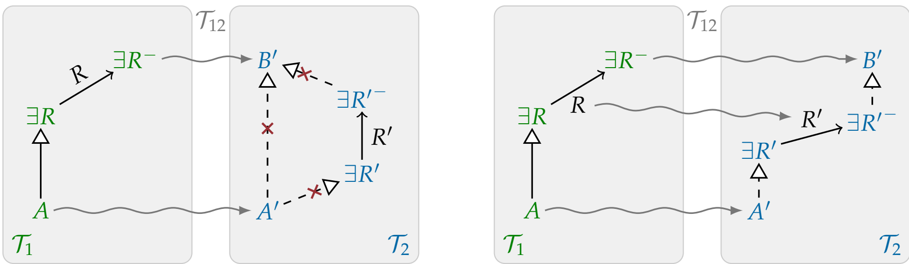
(1) $\mathcal{T}_{12} = \{A \sqsubseteq A', \exists R^- \sqsubseteq B'\}$.
Then there exists no UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$: take $\mathcal{A}_1 = \{A(a)\}$ and a Boolean target query $q = \exists x.\big(A'(a) \wedge B'(x)\big)$. Then q and only q evaluates to true over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$. Let us consider two target TBoxes $\mathcal{T}_2$ such that q also evaluates to true over $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$:

a) $\mathcal{T}_2 = \{A' \sqsubseteq B'\}$. Then for the query $q' = B'(a)$, $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models q'$, while $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \not\models q'$. Hence $\mathcal{T}_2$ is not a UCQ-representation.

b) $\mathcal{T}_2 = \{A' \sqsubseteq \exists R', \exists R'^- \sqsubseteq B'\}$. Then for the query $q' = \exists x.R'(a,x)$, $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models q'$, while $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \not\models q'$. Hence $\mathcal{T}_2$ is not a UCQ-representation.

(2) $\mathcal{T}_{12} = \{A \sqsubseteq A', \exists R^- \sqsubseteq B', R \sqsubseteq R'\}$.
Then $\mathcal{T}_2 = \{A' \sqsubseteq \exists R', \exists R'^- \sqsubseteq B'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

These cases can be depicted in the following ER diagrams:



(1) (2)

■

In the following examples we consider also disjointness assertions. Now we will, however, fix the mapping, and see how UCQ-representations depend on the source ABox.

Example 4.3.3. Assume that $\mathcal{M} = (\{A(\cdot), B(\cdot), C(\cdot)\}, \{A'(\cdot), B'(\cdot)\}, \mathcal{T}_{12})$, where $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq \neg A'\}$, and let

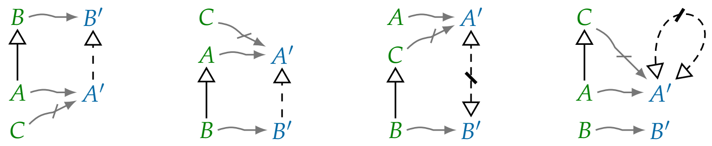
(1) $\mathcal{T}_1 = \{A \sqsubseteq B\}$.
Then TBox $\mathcal{T}_2 = \{A' \sqsubseteq B'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. First, notice that every source ABox $\mathcal{A}_1$ is consistent with $\mathcal{T}_1$. It should be clear that for every $\mathcal{A}_1 = \{X(a)\}$ for $X \in \{A, B, C\}$ or $\mathcal{A}_1 = \{B(a), C(a)\}$, $\mathcal{A}_1$ is consistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$, and $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = cert(q, \langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$ for each UCQ q.

Consider now $\mathcal{A}_1 = \{A(a), C(a)\}$, then $\mathcal{A}_1$ is not consistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$ (in fact, $\mathcal{A}_1$ is not consistent already with $\mathcal{T}_{12}$, so $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = AllTup(q)$ for each UCQ q. On the other hand, $\mathcal{A}_1$ is not consistent with $\mathcal{T}_2 \cup \mathcal{T}_{12}$ either, so as well, $cert(q, \langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = AllTup(q)$ for each UCQ q.

(2) $\mathcal{T}_1 = \{B \sqsubseteq A\}$.
Similarly to the previous case, TBox $\mathcal{T}_2 = \{B' \sqsubseteq A'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, but now it is a bit more involved. Namely, in this case ABox $\mathcal{A}_1 = \{B(a), C(a)\}$ is not consistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$, but consistent with $\mathcal{T}_{12}$ alone. But anyway, $\mathcal{A}_1$ is not consistent with $\mathcal{T}_2 \cup \mathcal{T}_{12}$ due to the axiom $B' \sqsubseteq A'$ in $\mathcal{T}_2$. So $cert(q, \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle) = cert(q, \langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$ for each ABox $\mathcal{A}_1$ and UCQ q.

(3) $\mathcal{T}_1 = \{B \sqsubseteq C\}$.
Then, TBox $\mathcal{T}_2 = \{B' \sqsubseteq \neg A'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. This case is in a way the opposite of (2). Consider ABox $\mathcal{A}_1 = \{A(a), B(a)\}$, then $\mathcal{A}_1$ is inconsistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$. Now, $\mathcal{A}_1$ is inconsistent with $\mathcal{T}_2 \cup \mathcal{T}_{12}$ is achieved with the disjointness assertion $B' \sqsubseteq A'$ in $\mathcal{T}_2$.

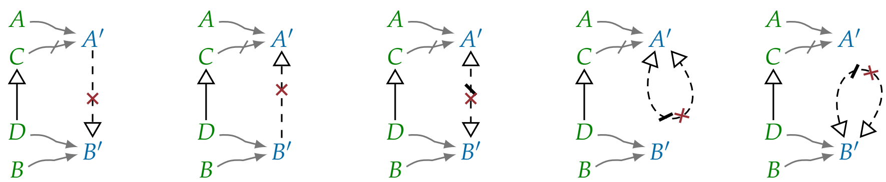
(4) $\mathcal{T}_1 = \{A \sqsubseteq C\}$.
Then, TBox $\mathcal{T}_2 = \{A' \sqsubseteq \neg A'\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. Observe, that every ABox $\mathcal{A}_1$ such that $A(a) \in \mathcal{A}_1$ for some constant a is inconsistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$. So the axiom $A' \sqsubseteq \neg A'$ in $\mathcal{T}_2$ assures that every such $\mathcal{A}_1$ is also inconsistent with $\mathcal{T}_2 \cup \mathcal{T}_{12}$. One the other hand, it is easy to see that for every source ABox that does not contain assertions with A, the required is satisfied.



(1) (2) (3) (4)

■

Example 4.3.4. Assume that $\mathcal{M} = (\{A(\cdot), B(\cdot), C(\cdot), D(\cdot)\}, \{A'(\cdot), B'(\cdot)\}, \mathcal{T}_{12})$, where $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq \neg A', D \sqsubseteq B'\}$, and let $\mathcal{T}_1 = \{D \sqsubseteq C\}$.
Then there exists no target TBox $\mathcal{T}_2$ that is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. In fact, it is easy to see that none of the TBoxes $\{A' \sqsubseteq B'\}$ or $\{B' \sqsubseteq A'\}$ is a UCQ-representation by taking ABoxes $\mathcal{A}_1 = \{A(a)\}$ or $\mathcal{A}_1 = \{B(a)\}$, respectively, to derive counterexamples. The next possibility is $\mathcal{T}_2 = \{A' \sqsubseteq \neg B'\}$, however if we consider source ABox $\mathcal{A}_1 = \{A(a), B(a)\}$, then $\mathcal{A}_1$ is consistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$, but inconsistent with $\mathcal{T}_2 \cup \mathcal{T}_{12}$, so for $q = A'(b)$ where b is a constant distinct from a, $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \not\models q$, and $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models q$. Finally, one can see that neither $\mathcal{T}_2 = \{A' \sqsubseteq \neg A'\}$ nor $\mathcal{T}_2 = \{B' \sqsubseteq \neg B'\}$ is a UCQ-representation by taking ABoxes $\mathcal{A}_1 = \{A(a)\}$ or $\mathcal{A}_1 = \{B(a)\}$, respectively. Below we depict the ER diagrams of $\mathcal{T}_1$, $\mathcal{T}_{12}$, and $\mathcal{T}_2$ for all possible target TBoxes $\mathcal{T}_2$:



■

5

REASONING ABOUT UNIVERSAL SOLUTIONS

In this chapter, we study the membership and non-emptiness problems for universal solutions, in the cases where nulls are not allowed (Section 5.2) and are allowed (Section 5.3) in such solutions, and conclude with the case of *DL-Lite*$_{RDFS}$ (Section 5.4). But before, in Section 5.1 we present a characterization of universal solutions in *DL-Lite*$_{\mathcal{R}}$.

5.1 CHARACTERIZATION OF UNIVERSAL SOLUTIONS

We start by defining the notion of Σ_2-safeness required to deal with disjointness assertions in the source KBs and mappings. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ is a mapping and $\mathcal{K}_1$ is a KB over Σ_1. Let $\mathcal{K} = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ and $\mathcal{U}_{\mathcal{K}}$ the canonical model of $\mathcal{K}$. Then a basic concept B over Σ_1 is said to be Σ_2-*safe in* $\mathcal{U}_{\mathcal{K}}$ if for every $d \in B^{\mathcal{U}_{\mathcal{K}}}$

$$d \notin N_a \quad \text{and} \quad \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d) = \emptyset.$$

Intuitively, safeness for B means no constant "associated" with B and no target concept "associated" with B via $\mathcal{T}_1$ and $\mathcal{T}_{12}$ can be mentioned in the target; in Example 4.1.7 neither A nor B is safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Furthermore, a pair of basic concepts (B, C) is is said to be Σ_2-*safe* in $\mathcal{U}_{\mathcal{K}}$ if B or C is safe. Intuitively, if a pair (B, C) is not safe and $(B \sqsubseteq \neg C) \in \mathcal{T}_1$, then universal solutions cannot exist, as explained in Example 4.1.7. Similarly, we say a basic role R over Σ_1 is Σ_2-*safe in* $\mathcal{U}_{\mathcal{K}}$ if for every $(d, d') \in R^{\mathcal{U}_{\mathcal{K}}}$

$$\begin{array}{llll} \text{either} & d \notin N_a & \text{and} & \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d) = \emptyset, \\ \text{or} & d' \notin N_a & \text{and} & \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d') = \emptyset. \end{array}$$

A pair of roles (R, Q) is Σ_2-*safe in* $\mathcal{U}_{\mathcal{K}}$ if 1) R or Q is safe, and 2) $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d') = \emptyset$ or $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(d'') = \emptyset$ for every $d, d', d'' \in \Delta^{\mathcal{U}_{\mathcal{K}}}$ such that $(d, d') \in R^{\mathcal{U}_{\mathcal{K}}}$ and $(d, d'') \in Q^{\mathcal{U}_{\mathcal{K}}}$.

Definition 5.1.1. *$\mathcal{K}_1$ is* Σ_2-safe *with respect to $\mathcal{M}$ if*

(CSAFE) *each pair of concepts (B, C) is Σ_2-safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ whenever $B \sqsubseteq \neg C \in \mathcal{T}_1$,*

(RSAFE) *each pair of roles (R, Q) is Σ_2-safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ whenever $R \sqsubseteq \neg Q \in \mathcal{T}_1$,*

(CEMPTY) *$B^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} = \emptyset$ for each basic concept B such that $B \sqsubseteq \neg B' \in \mathcal{T}_{12}$,*

(REMPTY) *$R^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} = \emptyset$ for each basic role R such that $R \sqsubseteq \neg R' \in \mathcal{T}_{12}$.*

Note that, if $\mathcal{K}_1$ and $\mathcal{M}$ do not contain disjointness assertions, then $\mathcal{K}_1$ is trivially Σ_2-safe with respect to $\mathcal{M}$.

5.1.1 *Characterization of the membership problem*

Now, we are ready to provide a characterization of universal solutions, where we already take into account Proposition 4.1.4, and therefore consider only target ABoxes as universal solutions. Observe that one of the conditions here is that a universal solution has to be homomorphically equivalent to the canonical model of the source KB and the mapping, which is similar to the characterization of universal solutions in the classical data exchange setting [51, 52].

Lemma 5.1.2. *An ABox $\mathcal{A}_2$ over Σ_2 is a universal solution for a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ iff the following conditions hold:*

(SAFE) *$\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$.*

(HOM) *$\mathcal{U}_{\mathcal{A}_2}$ is Σ_2- homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$,*

Proof. ($\Rightarrow$) Let $\mathcal{A}_2$ be a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Then $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$: since $\mathcal{A}_2$ is a solution, there exists $\mathcal{I}$, a model of $\mathcal{K}_1$, such that $(\mathcal{I}, \mathcal{U}_{\mathcal{A}_2}) \models \mathcal{T}_{12}$. Then $\mathcal{I} \cup \mathcal{U}_{\mathcal{A}_2}$ is a model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, therefore there is a homomorphism h from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{I} \cup \mathcal{U}_{\mathcal{A}_2}$. As Σ_1 and Σ_2 are disjoint signatures it follows that h is a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$. On the other hand, as $\mathcal{A}_2$ is a universal solution, $\mathcal{J}$, the interpretation of Σ_2 obtained from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is a model of $\mathcal{A}_2$ with a substitution h'. This h' is exactly a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Thus, we showed (hom).

For the sake of contradiction, assume that (safe) does not hold, i.e., $\mathcal{K}_1$ is not Σ_2-safe with respect to $\mathcal{M}$, and e.g., (csafe) does not hold, i.e., there is a disjointness constraint in $\mathcal{T}_1$ of the form $B \sqsubseteq \neg C$, such that (B, C) is not Σ_2-safe. Then both B and C are not Σ_2-safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$: for some $b \in B^{\mathcal{U}_\mathcal{K}}$ and $c \in C^{\mathcal{U}_\mathcal{K}}$,

$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b) \neq \emptyset \quad \text{or} \quad b \in \mathsf{N}_a, \qquad \text{and} \qquad \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c) \neq \emptyset \quad \text{or} \quad c \in \mathsf{N}_a.$$

Let h be a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$ (it exists by (hom)), and $h(b) = t$ and $h(c) = s$. Then it follows that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(t) \neq \emptyset \quad \text{or} \quad b \in \mathsf{N}_a, \qquad \text{and} \qquad \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(s) \neq \emptyset \quad \text{or} \quad c \in \mathsf{N}_a.$$

Take a model $\mathcal{J}$ of $\mathcal{A}_2$ with a substitution $h_\mathcal{J}$ such that $\Delta^\mathcal{J} = \{d\}$ (hence, $t^\mathcal{J} = s^\mathcal{J}$). Such a model exists because $\mathcal{A}_2$ does not assert any negative information and there is no UNA. First, assume that both b and c are constants (i.e., $b^\mathcal{J} = c^\mathcal{J}$). Then, obviously there exists no model $\mathcal{I}$ of Σ_1 such that $\mathcal{I} \models \mathcal{K}_1$ and $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$: in every such $\mathcal{I}$, $b^\mathcal{I}$ must be equal to $c^\mathcal{I}$ which contradicts $B \sqsubseteq \neg C$, and $b^\mathcal{I} \in B^\mathcal{I}$ and $c^\mathcal{I} \in C^\mathcal{I}$. Now, assume that at least b is not a constant and $\mathsf{tail}(b) = w_{[R]}$ for some role R over Σ_1 (hence, $b \in (\exists R^-)^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ and $\mathcal{T}_1 \models \exists R^- \sqsubseteq B$). Let $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b)$, then by construction of the canonical model, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists R^- \sqsubseteq B'$, by homomorphism, $B'(t) \in \mathcal{A}_2$, and by construction of $\mathcal{J}$, $B'^\mathcal{J} = \{d\}$. As $\mathcal{A}_2$ is a universal solution, let $\mathcal{I}$ be a model of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$. Then $(\exists R^-)^\mathcal{I}$ is non-empty and $(\exists R^-)^\mathcal{I} \subseteq B'^\mathcal{J}$. It immediately follows that $d \in (\exists R^-)^\mathcal{I}$, hence $d \in B^\mathcal{I}$. By a similar argument, it can be shown that d must be in $C^\mathcal{I}$, which contradicts that $\mathcal{I}$ is a model of $B \sqsubseteq \neg C$. Contradiction with $\mathcal{A}_2$ being a universal solution.

Similar to (csafe) we can derive a contradiction if assume that (rsafe) does not hold.

Now, assume (rempty) does not hold, i.e., $B \sqsubseteq \neg B' \in \mathcal{T}_{12}$ and $B^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} \neq \emptyset$. Note that $\mathcal{A}_2$ is an extended ABox, i.e., it contains only assertions of the form $A(u)$, $P(u,v)$ for $u, v \in N_a \cup N_l$. Take a model $\mathcal{J}$ of $\mathcal{A}_2$ such that $B'^{\mathcal{J}} = \Delta^{\mathcal{J}}$. Such $\mathcal{J}$ exists as $\mathcal{A}_2$ contains only positive facts. Since $\mathcal{A}_2$ is a universal solution, there exist a model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$. Then, $B^{\mathcal{I}} \neq \emptyset$, and it is easy to see that $(\mathcal{I}, \mathcal{J}) \not\models B \sqsubseteq \neg B'$ because $\Delta^{\mathcal{J}} \setminus B'^{\mathcal{J}} = \emptyset$ and $B^{\mathcal{I}} \not\subseteq \Delta^{\mathcal{J}} \setminus B'^{\mathcal{J}}$.

Similar to (cempty) we can derive a contradiction if assume that (rempty) does not hold. In every case we derive a contradiction, hence $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$.

($\Leftarrow$) Assume (hom) and (safe) hold. We show that $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$.

First, $\mathcal{A}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$. Let $\mathcal{J}$ be a model of $\mathcal{A}_2$, and h_1 a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{J}$. Furthermore, let h be a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$. Then $h_2(x) = h_1(h(x))$ is a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{J}$. Let $\mathcal{I}$ be the interpretation of Σ_1 defined as the image of h_2 applied to $\mathcal{U}_{\mathcal{K}_1}$, i.e., $\mathcal{I} = h_2(\mathcal{U}_{\mathcal{K}_1})$. Next, define a new function $h' : \Delta^{\mathcal{U}_{\mathcal{K}_1}} \to \Delta \cup \Delta^{\mathcal{I}}$, where Δ is an infinite set of domain elements disjoint from $\Delta^{\mathcal{I}}$, as follows:

- $h'(x) = h_2(x)$ if $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(x) \neq \emptyset$ or $x \in N_a$.
- $h'(x) = d_x$, a fresh domain element from Δ, otherwise.

We show that interpretation $\mathcal{I}'$ defined as the image of h' applied to $\mathcal{U}_{\mathcal{K}}$, is a model of $\mathcal{K}_1$ and $(\mathcal{I}', \mathcal{J}) \models \mathcal{M}$. It is straightforward to verify that $\mathcal{I}'$ is a model of the positive inclusions in $\mathcal{T}_1$ and $(\mathcal{I}', \mathcal{J})$ satisfy the positive inclusions from $\mathcal{T}_{12}$. In what follows we prove that $\mathcal{I}'$ is a model of the disjointness assertions in $\mathcal{T}_1$.

Let $\mathcal{T}_1 \models B \sqsubseteq \neg C$ for basic concepts B, C. By contradiction, assume $\mathcal{I}' \not\models B \sqsubseteq \neg C$, i.e., for some $d \in \Delta^{\mathcal{I}'}$, $d \in B^{\mathcal{I}'} \cap C^{\mathcal{I}'}$. We defined $\mathcal{I}'$ as the image of h' on $\mathcal{U}_{\mathcal{K}_1}$, hence there must exist $b, c \in \Delta^{\mathcal{U}_{\mathcal{K}_1}}$ such that $b \in B^{\mathcal{U}_{\mathcal{K}_1}}$, $c \in C^{\mathcal{U}_{\mathcal{K}_1}}$, and $h'(b) = h'(c) = d$. Then, since $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, it follows that (B, C) is Σ_2-safe and it cannot be the case that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b) \neq \emptyset \quad \text{or} \quad b \in N_a, \qquad \text{and} \qquad \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c) \neq \emptyset \quad \text{or} \quad c \in N_a.$$

Assume b is a null and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b) = \emptyset$. Then by definition of h', $h'(b) = d_b \in \Delta$ (hence $d = d_b$). In either case c is a constant, or $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c) \neq \emptyset$, or $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c) = \emptyset$, we obtain contradiction with $h'(b) = d_b = h'(c)$ (recall, Δ and $\Delta^{\mathcal{I}}$ are disjoint). Contradiction rises from the assumption $\mathcal{I} \not\models B \sqsubseteq \neg C$.

Next, assume $\mathcal{T}_1 \models R \sqsubseteq \neg Q$ for roles R, Q, and $\mathcal{I}' \not\models R \sqsubseteq \neg Q$, i.e., for some $d_1, d_2 \in \Delta^{\mathcal{I}'}$, $(d_1, d_2) \in R^{\mathcal{I}'} \cap Q^{\mathcal{I}'}$. We defined $\mathcal{I}'$ as the image of h' on $\mathcal{U}_{\mathcal{K}_1}$, hence there must exist $b_1, b_2, c_1, c_2 \in \Delta^{\mathcal{U}_{\mathcal{K}_1}}$ such that $(b_1, b_2) \in R^{\mathcal{U}_{\mathcal{K}_1}}$, $(c_1, c_2) \in Q^{\mathcal{U}_{\mathcal{K}_1}}$, and $h'(b_i) = h'(c_i) = d_i$ for $i = 1, 2$. Then, since $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, it follows that (R, Q) is Σ_2-safe and it cannot be the case that 1) R and Q are not Σ_2-safe, i.e.,

$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b_i) \neq \emptyset \quad \text{or} \quad b_i \in N_a, \qquad \text{and} \qquad \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c_i) \neq \emptyset \quad \text{or} \quad c_i \in N_a,$$

or 2) $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b_2) \neq \emptyset$ and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c_2) \neq \emptyset$ if $b_1 = c_1$. Consider the following possible cases:

- b_1 is a null and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b_1) = \emptyset$. Then by definition of h', $h'(b_1) = d_{b_1} \in \Delta$ (and $d_1 = d_{b_1}$).
  - c_1 is a null and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c_1) = \emptyset$, then $h'(c_1) = d_{c_1} = d_1$, hence $c_1 = b_1$ and $(b_1, b_2) \in R^{\mathcal{U}_{\mathcal{K}_1}}$, $(b_1, c_2) \in Q^{\mathcal{U}_{\mathcal{K}_1}}$. Assume b_2 is a null and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(b_2) = \emptyset$. Then $h'(b_2) = d_{b_2} \in \Delta$ and in either case c_2 is a constant, or $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c_2) \neq \emptyset$, or $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(c_2) = \emptyset$, we obtain contradiction with $h'(b_2) = d_{b_2} = h'(c_2)$.
  - otherwise we obtain contradiction with $h'(b_1) = d_{b_1} = h'(c_1)$.

The cases b_2 or c_i are nulls with the empty Σ_2-type are covered by swapping R and Q or by taking their inverses.

Finally, assume $B \sqsubseteq \neg B' \in \mathcal{T}_{12}$ and $(\mathcal{I}', \mathcal{J}) \not\models B \sqsubseteq \neg B'$, i.e., for some $d \in B^{\mathcal{I}'}$, $d \notin \Delta^{\mathcal{J}} \setminus C^{\mathcal{J}}$. Then there must exist $b \in B^{\mathcal{U}_{\mathcal{K}_1}}$ such that $h'(b) = d$. Contradiction with (cempty). Similarly, we derive a contradiction with (rempty) if assume that $R \sqsubseteq \neg R' \in \mathcal{T}_{12}$ and $(\mathcal{I}', \mathcal{J}) \not\models R \sqsubseteq \neg R'$.

Therefore, indeed, $\mathcal{I}$ is a model of $\mathcal{K}_1$ and $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$. This concludes the proof $\mathcal{A}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$.

Second, $\mathcal{A}_2$ is a universal solution. Let $\mathcal{I}$ be a model of $\mathcal{K}_1$ and $\mathcal{J}$ an interpretation of Σ_2 such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{M}$. Then, since $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is the canonical model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, there exists a homomorphism h from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{I} \cup \mathcal{J}$ ($\mathcal{I} \cup \mathcal{J}$ is a model of $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$). In turn, there is a homomorphism h_1 from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, therefore $h' = h \circ h_1$ is a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{I} \cup \mathcal{J}$, and a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{J}$. Hence, $\mathcal{J}$ is a model of $\mathcal{A}_2$: take h' as the substitution for the labeled nulls. By definition of universal solution, $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. □

The following examples illustrate why condition (safe) is needed in the characterization of universal solutions.

Example 5.1.3. Consider Example 4.1.7, i.e., $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot)\}$, $\Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{A \sqsubseteq A',\ B \sqsubseteq B'\}$, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{A \sqsubseteq \neg B\}$ and $\mathcal{A}_1 = \{A(a), B(b)\}$. We already argued that the target ABox $\mathcal{A}_2 = \{A'(a), B'(b)\}$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. In fact, condition (safe) is not satisfied because there is a disjointness assertion $A \sqsubseteq \neg B$ in $\mathcal{T}_1$ such that the pair (A, B) is not Σ_2-safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$: neither A, nor B is Σ_2-safe as $a \in A^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} \cap N_a$ and $b \in B^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} \cap N_a$.

We stress the attention of the reader that in the example above the disjointness assertion $A \sqsubseteq \neg B$ forces that a and b be interpreted as different elements in $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$. Without the UNA there are no means to express that in the target (the language of target ABoxes is that one of extended ABoxes, which can express only positive information), that is why $\mathcal{A}_2$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. However, ABoxes that allow for inequality between constants would do the job: thus, the ABox with inequalities $\{A'(a), B'(b), a \neq b\}$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

Example 5.1.4. Assume $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{R(\cdot,\cdot), S(\cdot,\cdot)\}$, $\Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{\exists R^- \sqsubseteq A',\ \exists S \sqsubseteq B'\}$, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{\exists R^- \sqsubseteq \neg \exists S^-\}$ and $\mathcal{A}_1 = \{\exists R(a), \exists S(a)\}$. Then ABox $\mathcal{A}_2 = \{A'(n), B'(m)\}$

is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Again, there is a disjointness assertion $\exists R^- \sqsubseteq \neg\exists S^-$ in $\mathcal{T}_1$ such that the pair $(\exists R^-, \exists S^-)$ is not Σ_2-safe in $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$: $\exists R^-$ is not Σ_2-safe because $aw_{[R]} \in (\exists R^-)^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(aw_{[R]}) = \{A'\}$, and analogously for $\exists S^-$.

Observe that in this example for an ABox $\mathcal{A}_2$ to be a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, one has to enforce that the labeled nulls corresponding to $aw_{[R]}$ and $aw_{[S]}$ be interpreted as distinct elements, which requires inequality between labeled nulls in the language. ∎

Example 5.1.5. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot)\}, \Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq \neg B'\}$, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{T}_1 = \{\}$ and $\mathcal{A}_1 = \{A(a), B(b)\}$. Then ABox $\mathcal{A}_2 = \{A'(a)\}$ is not a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Namely, condition (cempty) is violated: there is a disjointness assertion $B \sqsubseteq \neg B'$ in $\mathcal{T}_{12}$ such that $B^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ is non-empty (to be precise $B^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} = \{b\}$). The latter implies that an interpretation $\mathcal{J}$ of Σ_2 such that $b^{\mathcal{J}} \in B'^{\mathcal{J}}$ cannot be a model of $\mathcal{A}_2$, but obviously it is not the case here.

In this example, one needs to consider ABoxes with negative atoms to obtain a universal solution: the ABox $\{A'(a), \neg B'(b)\}$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Since we require target ABoxes to be extended ABoxes, no extended ABox can be a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ∎

Hence, we have seen in the three examples above that the main reason for condition (safe) in the characterization of universal solutions is the inability of extended ABoxes to express any form of negative information, be it inequality between terms or negation between atoms.

Condition (safe) in Lemma 5.1.2 is easy: it can be checked in NLOGSPACE using an algorithm, based on directed graph reachability solving procedure. Given a KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ the graph $\mathsf{G}_{\mathcal{K}}$ is defined as essentially the ER diagram of $\mathcal{T}$ plus the edges from a constant a to a concept B if $\mathcal{A} \models B(a)$. Then for each concept B and each role R we can check in NLOGSPACE whether, respectively, $B^{\mathcal{U}_{\mathcal{K}}}$ and $R^{\mathcal{U}_{\mathcal{K}}}$ is non-empty by verifying that B and R, respectively, is reachable in $\mathsf{G}_{\mathcal{K}}$ from some constants; for each concept B we can check in NLOGSPACE whether $B^{\mathcal{U}_{\mathcal{K}}}$ contains any constants or any element of the form $\sigma w_{[S]}$ such that it is possible to reach in $\mathsf{G}_{\mathcal{K}}$ from the concept $\exists S^-$ some target concept (that is, $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(\sigma w_{[S]}) \neq \emptyset$).

As for condition (hom), we show how to check it in Section 5.2 for simple universal solutions, i.e., when we consider only simple target ABoxes, and in Section 5.3 for extended universal solutions, i.e., when we consider extended target ABoxes.

5.1.2 *Characterization of the non-emptiness problem*

Next, we provide a characterization of the cases when a universal solution exists. Again, this result is related to the problem of existence of a finite "core" in the classical data exchange setting [52, 62].

Lemma 5.1.6. *Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a mapping, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ a KB over Σ_1. Then, a universal solution with extended ABoxes for $\mathcal{K}_1$ under $\mathcal{M}$ exists iff*

(SAFE) *$\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$.*

(CORE) *$\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$ is Σ_2-homomorphically embeddable into a finite subset of itself.*

Proof. ($\Leftarrow$) Let ABox $\mathcal{A}_2$ be an ABox over Σ_2 such that $\mathcal{U}_{\mathcal{A}_2}$ is a finite subset of $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$ and there exists a Σ_2-homomorphism h from $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$. Then, $\mathcal{U}_{\mathcal{A}_2}$ is trivially homomorphically embeddable into $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$. Since, $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, by Lemma 5.1.2, we obtain that $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$.

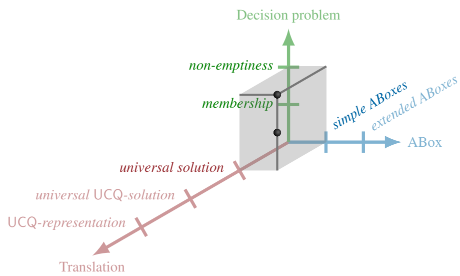
($\Rightarrow$) Let $\mathcal{A}_2$ be a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. First, it immediately follows that $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$. Then, $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$ by Lemma 5.1.2. Let h be a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$, and $h(\mathcal{U}_{\mathcal{A}_2})$ the image of h. Then, $h(\mathcal{U}_{\mathcal{A}_2})$ is a finite subset of $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$, moreover it is homomorphically equivalent to $\mathcal{U}_{\mathcal{A}_2}$ and to $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$. Therefore, it follows that $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$ is Σ_2-homomorphically embeddable to a finite subset of itself. $\square$

It follows from the proof of Lemma 5.1.6 that the ABox $\mathcal{A}_2$ corresponding to the finite subset $\mathcal{U}_{\mathcal{A}_2}$ of $\mathcal{U}_{\langle\mathcal{T}_1\cup\mathcal{T}_{12},\mathcal{A}_1\rangle}$ in condition (core) is a universal solution. Hence, if we additionally require in condition (core) that the finite subset $\mathcal{U}_{\mathcal{A}_2}$ does not contain anonymous individuals, we obtain a characterization for universal solutions with simple ABoxes.

We introduce some additional notation required for this chapter. Let $\mathcal{K}$ be an OWL 2 QL KB. Given a subset D of $\Delta^{\mathcal{U}_\mathcal{K}}$, the sub-interpretation of $\mathcal{U}_\mathcal{K}$ *induced by* D, denoted by $\mathcal{U}_\mathcal{K}^D$, is defined as $\Delta^{\mathcal{U}_\mathcal{K}^D} = D$, $A^{\mathcal{U}_\mathcal{K}^D} = A^{\mathcal{U}_\mathcal{K}} \cap D$ for each concept name A, and $P^{\mathcal{U}_\mathcal{K}^D} = P^{\mathcal{U}_\mathcal{K}} \cap (D \times D)\}$ for each role name P. Furthermore, let $a \in \mathsf{Ind}(\mathcal{K})$, and D^a the set of elements in $\mathsf{path}(\mathcal{K})$ beginning with a, i.e., $D^a = \{a\sigma \mid a\sigma \in \Delta^{\mathcal{U}_\mathcal{K}}\}$. Then the sub-interpretation of $\mathcal{U}_\mathcal{K}$ *generated by* a, denoted by $\mathcal{U}_\mathcal{K}^a$, is defined as $\Delta^{\mathcal{U}_\mathcal{K}^a} = D^a$, $A^{\mathcal{U}_\mathcal{K}^a} = A^{\mathcal{U}_\mathcal{K}} \cap D^a$ for each concept name A, and $P^{\mathcal{U}_{\mathcal{K}_1}^a} = (P^{\mathcal{U}_{\mathcal{K}_1}} \cap (D^a \times D^a)) \setminus \{(a,a)\}$ for each role name P. Notice that $\mathcal{U}_\mathcal{K}^a$ describes a tree structure. $\mathcal{G}_\mathcal{K}^D$ and $\mathcal{G}_\mathcal{K}^a$ are defined accordingly.

5.2 SIMPLE UNIVERSAL SOLUTIONS

In this section, we show that both the membership and the non-emptiness problems for universal solutions without null values are PTIME-complete. The upper bound is obtained by a novel reduction to reachability games on graphs.



5.2.1 *The non-emptiness problem*

We start with tackling the non-emptiness problem, and first provide a PTIME lower bound for this problem.

Lemma 5.2.1. *The non-emptiness problem for simple universal solutions is* PTIME-*hard in data complexity.*

Proof. The proof is by reduction from the Circuit Value Problem known to be PTIME-hard: given a monotone Boolean circuit C consisting of assignments of the form $P_i = 0$, $P_i = 1$, $P_i = P_j \wedge P_k$, $j, k < i$, or $P_i = P_j \vee P_k$, $j, k < i$, where each P_i appears on the left-hand side of exactly one assignment, compute the value of P_n.

We fix signatures $\Sigma_1 = \{P(\cdot), L(\cdot,\cdot), R(\cdot,\cdot)\}$ and $\Sigma_2 = \{L'(\cdot,\cdot), R'(\cdot,\cdot)\}$. Let $a_1, \dots, a_n \in N_a$, and consider

$$\begin{aligned}
\mathcal{A}_1 &= \{P(a_n) \cup \{L(a_i, a_i), R(a_i, a_i) \mid P_i = 1 \text{ in } C\} \\
&\quad \cup \{L(a_i, a_j), R(a_i, a_k) \mid P_i = P_j \wedge P_k \text{ in } C\} \\
&\quad \cup \{L(a_i, a_j), R(a_i, a_j), L(a_i, a_k), R(a_i, a_k) \mid P_i = P_j \vee P_k \text{ in } C\} \\
\mathcal{T}_1 &= \{P \sqsubseteq \exists L,\ P \sqsubseteq \exists R,\ \exists L^- \sqsubseteq P,\ \exists R^- \sqsubseteq P\}, \\
\mathcal{T}_{12} &= \{L \sqsubseteq L',\ R \sqsubseteq R'\}
\end{aligned}$$

Note that $\mathcal{T}_1$ and $\mathcal{M}$ do not depend on C, hence the reduction provide the lower bound for data complexity. We show that the value of P_n in C is true if and only if there exists a (simple) universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$.

($\Rightarrow$) Suppose P_n evaluates to true in C. We verify that $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically embeddable into $\mathcal{U}^{\mathsf{Ind}(\mathcal{A}_1)}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, or equivalently into $\mathcal{U}_{\mathcal{A}_2}$, where

$$\mathcal{A}_2 = \{L'(a_i, a_j) \mid L(a_i, a_j) \in \mathcal{A}_1\} \cup \{R'(a_i, a_j) \mid R(a_i, a_j) \in \mathcal{A}_1\}.$$

Observe that $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ contains an infinite binary tree with the root in a_n, one edge labeled with L' and the other edge labeled with R'. By induction, we define a Σ_2-homomorphism h from this tree to $\mathcal{U}_{\mathcal{A}_2}$.

First, it should be clear that $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(a_n) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(a_n)$, so we set $h(a_n) = a_n$.

Assume the value of P_i is true and we already defined $h(\sigma) = a_i$ for $\sigma \in \Delta^{\mathcal{U}^{a_n}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$. Consider the following three cases. First, $P_i = P_j \wedge P_k$ in C, then $\mathcal{A}_2$ contains assertions $L'(a_i, a_j)$ and $R'(a_i, a_k)$, moreover, P_j and P_k both evaluate to true: we set $h(\sigma w_{[L]}) = a_j$ and $h(\sigma w_{[R]}) = a_k$. Second, $P_i = P_j \vee P_k$ in C, then $\mathcal{A}_2$ contains assertions $L'(a_i, a_j), R'(a_i, a_j)$ and $L'(a_i, a_k), R'(a_i, a_k)$, and at least one of P_j and P_k evaluates to true, assume P_j: we set $h(\sigma w_{[L]}) = a_j$ and $h(\sigma w_{[R]}) = a_j$. Finally, if $P_i = 1$ in C, then $\mathcal{A}_2$ contains assertions $L'(a_i, a_i)$ and $R'(a_i, a_j)$: we set $h(\sigma w_{[L]}) = a_i$ and $h(\sigma w_{[R]}) = a_i$.

($\Leftarrow$) Suppose $\mathcal{A}_2$ is a simple target ABox that is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Then $\mathcal{A}_2$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, and it follows

$$\mathcal{A}_2 = \{L'(a_i, a_j) \mid L(a_i, a_j) \in \mathcal{A}_1\} \cup \{R'(a_i, a_j) \mid R(a_i, a_j) \in \mathcal{A}_1\}.$$

We prove that the value of P_n is true in C.

Let h be a Σ-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$. Since $\mathcal{U}^{a_n}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is an infinite tree, and the only role cycles that $\mathcal{A}_2$ contains are loops of the form $L'(a_i, a_i)$ and $R'(a_i, a_i)$, there exists a bound m such that for each $\sigma = a_n w_{[S_1]} \cdots w_{[S_m]} \in \Delta^{\mathcal{U}^{a_n}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ with $S_j \in \{L, R\}$, it holds $h(\sigma) = a_i$ for some i such that $P_i = 1$ in C. Assume $l > 1$ and the value of P_i is true in C whenever $h(a_n w_{[S_1]} \cdots w_{[S_l]}) = a_i$, for each $a_n w_{[S_1]} \cdots w_{[S_l]}$ with $S_j \in \{L, R\}$ and each $1 \le i < n$. We verify that the value of P_i is true in C whenever $h(a_n w_{[S_1]} \cdots w_{[S_{l-1}]}) = a_i$, for each $a_n w_{[S_1]} \cdots w_{[S_{l-1}]}$ and each $1 \le i \le n$.

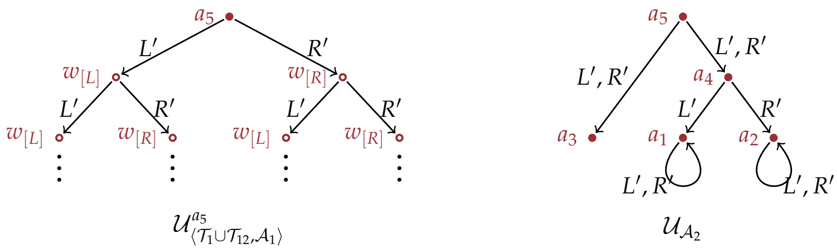
Assume that $h(a_n w_{[S_1]} \cdots w_{[S_{l-1}]} w_{[L]}) = a_j$, $h(a_n w_{[S_1]} \cdots w_{[S_{l-1}]} w_{[R]}) = a_k$ and the values of P_j and P_k are true in C, moreover $h(a_n w_{[S_1]} \cdots w_{[S_{l-1}]}) = a_i$. If $i = j = k$, then obviously the value of P_i is true in C. Otherwise $i \neq j$ and $i \neq k$. If $j = k$, then given that h is a homomorphism, $\mathcal{A}_2$ contains assertions $L'(a_i, a_j)$ and $R'(a_i, a_j)$ (hence, $\mathcal{A}_1$ contains assertions $L(a_i, a_j)$ and $R(a_i, a_j)$). By construction of $\mathcal{A}_1$, it follows there is an assignment $P_i = P_j \vee P_{j'}$ in C for some j', as P_j is true, we obtain that P_i evaluates to true. If $j \neq k$, then $\mathcal{A}_2$ contains assertions $L'(a_i, a_j)$ and $R'(a_i, a_k)$, so by construction of $\mathcal{A}_1$ there is an assignment $P_i = P_j \wedge P_k$ or $P_i = P_j \vee P_k$ in C. Again it follows P_i evaluates to true in C.

By induction, it follows that P_n evaluates to true in C. □

Example 5.2.2. For a circuit C containing five assignments

$$P_1 = 1,\ P_2 = 1,\ P_3 = 0,\ P_4 = P_1 \wedge P_2, \text{ and } P_5 = P_3 \vee P_4,$$

the projections over Σ_2 of $\mathcal{U}^{a_5}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ and $\mathcal{U}_{\mathcal{A}_2}$ can be depicted as follows:



■

Next, instead of providing an upper bound, we reduce the non-emptiness problem for simple universal solutions to the corresponding membership problem that we show how to solve in the next section.

Lemma 5.2.3. *The non-emptiness problem for simple universal solutions is reducible to the membership problem for simple universal solutions in polynomial time.*

Proof. Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a source KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, and the question to answer is whether there exists a simple ABox over Σ_2 that is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Let $\mathcal{A}_2$ be the following target ABox: *(i)* $\mathsf{Ind}(\mathcal{A}_2) \subseteq \mathsf{Ind}(\mathcal{A}_1)$, *(ii)* $\mathcal{A}_2 \models B(a)$ iff $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models B(a)$, and *(iii)* $\mathcal{A}_2 \models R(a,b)$ iff $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models R(a,b)$, for all $a, b \in \mathsf{Ind}(\mathcal{A}_1)$, concept B and role R over Σ_2. We show that there exists a simple universal solution for $\mathcal{K}_1$ under $\mathcal{M}$ if and only if $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Direction ($\Leftarrow$) is trivial, below we prove ($\Rightarrow$).

Let $\mathcal{A}$ be a simple universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Then $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, and $\mathcal{U}_{\mathcal{A}}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. The latter implies $\mathsf{Ind}(\mathcal{A}) \subseteq \mathsf{Ind}(\mathcal{A}_1)$, and $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(a) = \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}}}(a)$, $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}(a,b) = \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}}}(a,b)$, for all $a, b \in \mathsf{Ind}(\mathcal{A}_1)$. Then it must be the case $\mathcal{A}$ coincides with $\mathcal{A}_2$, so $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. □

From the proof above it follows that if a universal solution exists, then it can be computed as the ABox $\mathcal{A}_2$ defined in the proof.

5.2.2 *The membership problem*

In this section we provide an upper bound for the membership problem for simple universal solutions.

Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a source KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, and a simple target ABox $\mathcal{A}_2$, and want to decide whether $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. According to Lemma 5.1.2, it is sufficient to check conditions (safe) and (hom). The former condition does not depend on $\mathcal{A}_2$ and as discussed above, can be checked in polynomial time. As for the latter condition, denote the KB $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ by $\mathcal{K}$. First, the existence of a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ for a simple ABox $\mathcal{A}_2$ amounts to checking,

$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(a) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(a) \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(a,b) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(a,b) \text{ for all } a,b \in \mathsf{Ind}(\mathcal{A}_2).$$

Second, as for the existence of a Σ_2-homomorphism in the opposite direction, it should be the case that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(a) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(a) \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}}}(a,b) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{A}_2}}(a,b) \text{ for all } a,b \in \mathsf{Ind}(\mathcal{A}_1).$$

Clearly, these two conditions can be checked in PTIME. Now, for $c \in \mathsf{Ind}(\mathcal{A}_1)$, to check how the tree $\mathcal{U}_{\mathcal{K}}^c$ can be Σ_2-homomorphically mapped to $\mathcal{U}_{\mathcal{A}_2}$, we are going to employ the technique of infinite games on graphs. Specifically, we show how this problem can be reduced to the problem of existence of a winning strategy for Duplicator in a reachability game, known to be solvable in polynomial time. For a short introduction to (reachability) games, we refer to Section A.1.2. Below we show how to construct the game $G_\Sigma^c(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$ for a KB $\mathcal{K}$, an ABox $\mathcal{A}$, a signature Σ, and $c \in \mathsf{Ind}(\mathcal{K})$.

REACHABILITY GAME $G_\Sigma^c(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$ is a pair (G_c, F_c), where G_c is the *game graph*, and F_c is the winning condition, that is, the set of states that Spoiler wants to reach.

The game graph $\mathsf{G}_c = (\mathsf{S}, \mathsf{D}, \mathsf{T})$ has the set of states of the kind $(u \mapsto a)$ and $(a,\ u \rightsquigarrow u')$, where $u, u' \in \Delta^{\mathcal{G}_\mathcal{K}^c}$ and $a \in \mathsf{Ind}(\mathcal{A})$:

- S consists of the states $(u \mapsto a)$ with $\mathbf{t}_\Sigma^{\mathcal{G}_\mathcal{K}}(u) \subseteq \mathbf{t}_\Sigma^{\mathcal{U}_\mathcal{A}}(a)$; intuitively, such states represent a mapping of $\delta \in \Delta^{\mathcal{U}_\mathcal{K}}$ with $\mathsf{tail}(\delta) = u$ to a. Given this partial homomorphism, Spoiler can decide to challenge Duplicator with one of the successors u' of u in $\mathcal{G}_\mathcal{K}^a$.
- D consists of the states $(a,\ u \rightsquigarrow u')$ with $u \rightsquigarrow_\mathcal{K} u'$; the states represent "challenges" that Duplicator must address by finding a constant $a' \in \mathsf{Ind}(\mathcal{A})$ so that the "challenged" edge $(\delta,\ \delta \cdot u')$ of the tree $\mathcal{U}_\mathcal{K}^a$ can be "mapped" to the edge (a, a') of $\mathcal{U}_\mathcal{A}$.

Therefore, the transitions between S and D, forming T, are defined as the union of:

- $\big((u \mapsto a),\ (a,\ u \rightsquigarrow u')\big)$, and
- $\big((a,\ u \rightsquigarrow u'),\ (u' \mapsto a')\big)$ whenever $\mathbf{r}_\Sigma^{\mathcal{G}_\mathcal{K}}(u,u') \subseteq \mathbf{r}_\Sigma^{\mathcal{U}_\mathcal{A}}(a,a')$.

Notice that the size of G_c is $O(|\mathcal{G}_\mathcal{K}^c| \times |\mathcal{A}|)$.

The set F_c, which is the set of states that Spoiler wants to reach, is given by the duplicating states that are "dead ends", i.e.,

$$F_c = \big\{(c,\ u \rightsquigarrow u') \mid (u' \mapsto a') \notin \mathsf{S} \text{ or } \mathbf{r}_\Sigma^{\mathcal{G}_\mathcal{K}}(u,u') \not\subseteq \mathbf{r}_\Sigma^{\mathcal{U}_\mathcal{A}}(c,c'), \text{for all } a' \in \Delta^{\mathcal{U}_\mathcal{A}}\big\}.$$

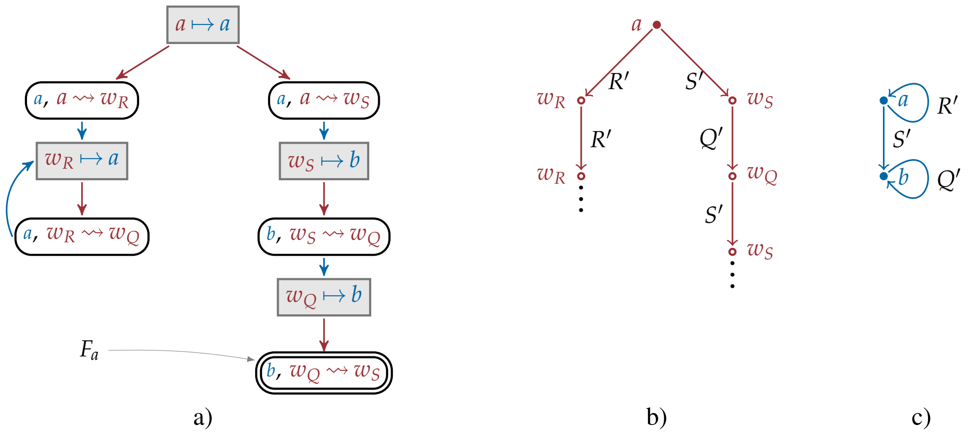


Figure 7: Example of a game: a) the game graph G_a, b) the projection of $\mathcal{U}_\mathcal{K}$ over Σ, c) $\mathcal{U}_\mathcal{A}$.

Intuitively, the game proceeds as follows. Duplicator tries to construct a Σ-homomorphism from the tree $\mathcal{U}^c_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$, and Spoiler attempts to fail him by finding a path in $\mathcal{U}^c_\mathcal{K}$ that does not have a homomorphic image in $\mathcal{U}_\mathcal{A}$, given the partial homomorphism constructed so far. Thus, Spoiler starts in $(u_0 \mapsto a_0)$ for $u_0 = a_0 = c$ if $(c \mapsto c) \in \mathsf{S}$, which corresponds to mapping c to c, and in each his turn chooses a successor u_{i+1} of u_i in $\mathcal{G}_\mathcal{K}$: the "challenge" represented by the state $(a_i,\ u_i \rightsquigarrow u_{i+1})$. Then Duplicator tries to find a constant c_{i+1} in $\mathsf{Ind}(\mathcal{A})$ that could be the image of the "challenged" element $u_0 \cdots u_{i+1}$ of $\mathcal{U}^c_\mathcal{K}$, i.e., he chooses a state $(u_{i+1} \mapsto a_{i+1})$ such that $\mathbf{r}^{\mathcal{G}_\mathcal{K}}_\Sigma(u_i, u_{i+1}) \subseteq \mathbf{r}^{\mathcal{U}_\mathcal{A}}_\Sigma(a_i, a_{i+1})$. Note that, if $\mathbf{r}^{\mathcal{G}_\mathcal{K}}_\Sigma(u_i, u_{i+1})$ is empty, then Duplicator can respond with any a_{i+1} such that $(u_{i+1} \mapsto a_{i+1})$ is a Spoiler state, even if a_{i+1} is not connected to a_i in $\mathcal{U}_\mathcal{A}$. Duplicator loses if at some point he cannot find where to map the challenged element, thus the game reaches a dead-end of Duplicator: $(a_i,\ u_i \rightsquigarrow u_{i+1}) \in F_c$. Otherwise, the game can reach a dead-end of Spoiler, or continue forever avoiding the dead-ends of Duplicator, hence Duplicator wins. Note that, if $(c \mapsto c) \notin \mathsf{S}$, then we assume that Spoiler "wins" the game immediately.

We illustrate such games in the following example.

Example 5.2.4. Assume $\Sigma = \{R'(\cdot,\cdot), S'(\cdot,\cdot), Q'(\cdot,\cdot)\}$, $\mathcal{K} = \langle \mathcal{T}, \{\exists R(a), \exists S(a)\}\rangle$, where $\mathcal{T} = \{\exists R^- \sqsubseteq \exists R, \exists S^- \sqsubseteq \exists Q, \exists Q^- \sqsubseteq \exists S, R \sqsubseteq R', S \sqsubseteq S', Q \sqsubseteq Q'\}$, and $\mathcal{A} = \{R'(a,a), S'(a,b), Q'(b,b)\}$. Then $G^a_\Sigma(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A}) = (\mathsf{G}_a, F_a)$, where $F_a = \{(b,\ w_Q \rightsquigarrow w_S)\}$, and the game graph G_a can be depicted as in a) in Figure 7 (we ignore the states that are not reachable from $(a \mapsto a)$; the duplicating states forming D are shown as ovals and the spoiling states forming S are shown as boxes). In b) we show the projection of $\mathcal{U}_\mathcal{K}$ over Σ, and in c) we show $\mathcal{U}_\mathcal{A}$.

The game starts in the state $(a \mapsto a)$. It corresponds to setting the homomorphic image of $a \in \Delta^{\mathcal{U}_\mathcal{K}}$ to $a \in \Delta^{\mathcal{U}_\mathcal{A}}$. Then Spoiler can choose one of the two successors of a in $\mathcal{U}_\mathcal{K}$: either aw_R or aw_S. If he chooses aw_R, it means he moves to the state $(a,\ a \rightsquigarrow w_R)$. Now, Duplicator has to respond by finding where in $\mathcal{U}_\mathcal{A}$ to map aw_R: he can map it only to a (note the role labels), so he moves to $(w_R \mapsto a)$. In this manner, they continue forever moving between the states $(a,\ w_R \rightsquigarrow w_R)$ and $(w_R \mapsto a)$, which corresponds to mapping paths of the form $aw_R \cdots w_R \in \Delta^{\mathcal{U}_\mathcal{K}}$ to $a \in \Delta^{\mathcal{U}_\mathcal{A}}$. Thus, this play is an infinite play:

$$(a \mapsto a), (a,\ a \rightsquigarrow w_R), (w_R \mapsto a), (a,\ w_R \rightsquigarrow w_R), (w_R \mapsto a), \ldots$$

and it is a win for Duplicator.

However, if Spoiler at his first move chooses the successor aw_S of a, hence moves to the state $(a,\, a \rightsquigarrow w_S)$, the game would finish soon in a dead-end of Duplicator: Duplicator finds a homomorphic image of $aw_S \in \Delta^{\mathcal{U}_\mathcal{K}}$ as $b \in \Delta^{\mathcal{U}_\mathcal{A}}$, then Spoiler picks the successor aw_Sw_Q of aw_S, and Duplicator sets the homomorphic image of aw_Sw_Q to b. Finally, when Spoiler chooses the next successor, $aw_Sw_Qw_S$ in this case, Duplicator fails to find where to map it, so the game ends in the dead-end of Duplicator. Therefore, the second play is a finite play:

$$(a \mapsto a), (a,\, a \rightsquigarrow w_S), (w_S \mapsto b), (b,\, w_S \rightsquigarrow w_Q), (w_Q \mapsto b), (b,\, w_Q \rightsquigarrow w_S)$$

and it is a win for Spoiler as the game reached a state from F_a. ∎

Having constructed the game $G^c_\Sigma(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A}) = (\mathsf{G}_c, F_c)$, we prove that verifying whether $\mathcal{U}_\mathcal{K}$ can be Σ-homomorphically mapped to $\mathcal{U}_\mathcal{A}$ reduces to checking whether both $(c \mapsto c)$ is a state in the game graph G_c (i.e., $\mathbf{t}^{\mathcal{G}_\mathcal{K}}_\Sigma(c) \subseteq \mathbf{t}^{\mathcal{U}_\mathcal{A}}_\Sigma(c)$) and Duplicator has a winning strategy in $G^c_\Sigma(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$ from $(c \mapsto c)$.

Lemma 5.2.5. *Let $\mathcal{K}$ be a KB, $\mathcal{A}$ an ABox and Σ a signature. There exists a Σ-homomorphism from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$ iff*

(ABOX) $\mathbf{r}^{\mathcal{U}_\mathcal{K}}_\Sigma(a,b) \subseteq \mathbf{r}^{\mathcal{U}_\mathcal{A}}_\Sigma(a,b)$ *for all $a, b \in \mathsf{Ind}(\mathcal{K})$;*

(WIN) *$(c \mapsto c)$ is a state in G_c and Duplicator has a winning strategy in $G^c_\Sigma(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A}) = (\mathsf{G}_c, F_c)$ from $(c \mapsto c)$, for each $c \in \mathsf{Ind}(\mathcal{K})$.*

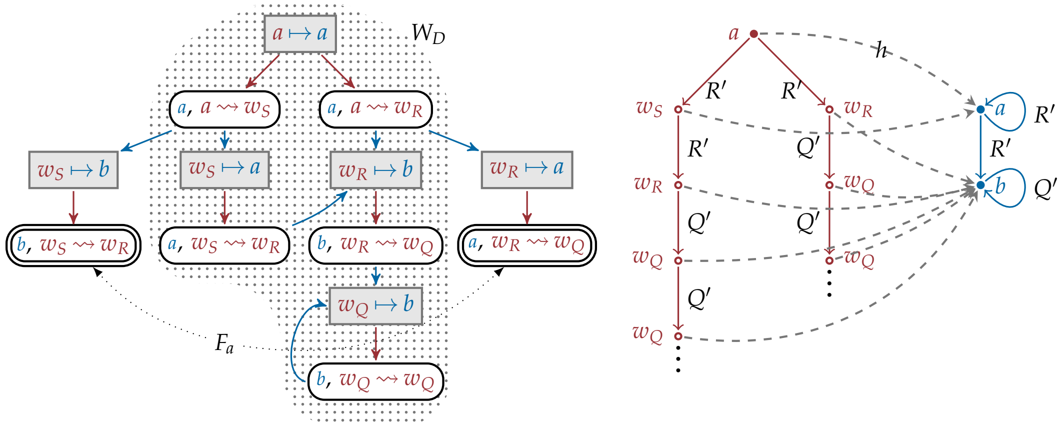
Proof. ($\Rightarrow$) Suppose h is a Σ_2-homomorphism from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$. Then clearly, (abox) and $\mathbf{t}^{\mathcal{G}_\mathcal{K}}_\Sigma(a) \subseteq \mathbf{t}^{\mathcal{U}_\mathcal{A}}_\Sigma(a)$ for each $a \in \mathsf{Ind}(\mathcal{K})$ hold. Let $c \in \mathsf{Ind}(\mathcal{K})$, and consider the game $G^c_\Sigma(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$. We describe a winning strategy f_1 for Duplicator in $G^a_\Sigma(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$ from $(c \mapsto c)$. Let

$$\pi = (u_0 \mapsto a_0), (a_0,\, u_0 \rightsquigarrow u_1), \ldots, (u_k \mapsto a_k), (a_k,\, u_k \rightsquigarrow u_{k+1})$$

be a play in G_a conform with f_1, where $k \geq 0$, $u_0 = a_0 = a$, and $a_i \in \mathsf{Ind}(\mathcal{A})$, $u_i \in \Delta^{\mathcal{G}^a_\mathcal{K}}$ for $i > 1$. Then we set $f_1(\pi) = (u_{k+1} \mapsto h(au_1 \cdots u_{k+1}))$. Note that by construction of T, $au_1 \cdots u_{k+1}$ is a valid path in $\mathcal{U}_\mathcal{K}$. Since h is defined for $\Delta^{\mathcal{U}_\mathcal{K}}$, it follows that f_1 is defined for each possible move of Spoiler, moreover, f_1 never assigns a dead-end of Duplicator to the current play π. Hence either the game ends in a dead-end of Spoiler (i.e., Spoiler is in a leaf of the tree in $\mathcal{U}_\mathcal{K}$), or it continues infinitely long avoiding visits to the dead-ends of Duplicator, in any case Duplicator wins.

($\Leftarrow$) Assume that both (abox) and (win) hold (in particular, $\mathbf{t}^{\mathcal{G}_\mathcal{K}}_\Sigma(a) \subseteq \mathbf{t}^{\mathcal{U}_\mathcal{A}}_\Sigma(a)$, for each $a \in \mathsf{Ind}(\mathcal{K})$). Given $c \in \mathsf{Ind}(\mathcal{K})$, we construct a Σ-homomorphism h_c from the tree $\mathcal{U}^c_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$. Let f be a winning strategy of Duplicator from $(c \mapsto c)$, and $\pi = (u_0 \mapsto a_0), (a_0,\, u_0 \rightsquigarrow u_1), \ldots, (u_k \mapsto a_k), (a_k,\, u_k \rightsquigarrow u_{k+1}), \ldots$ a play conforming with f, where $u_0 = a_0 = c$. Then Duplicator wins π, and either

- $\pi = (u_0 \mapsto a_0), (a_0,\, u_0 \rightsquigarrow u_1), \ldots, (u_k \mapsto a_k)$ is a *finite* play, $k \geq 0$, and $(u_k \mapsto a_k)$ is a dead-end of Spoiler. So we set $h_c(cu_1 \cdots u_i) = a_i$, for $0 \leq i \leq k$.
- π is an *infinite* play s.t. no states from F_c occur in it. We set $h_c(cu_1 \cdots u_i) = a_i$, for $i \geq 0$.



a) b)

Figure 8: Reduction: a) the game graph G_a, b) Σ-homomorphism h from $\mathcal{U}_{\mathcal{K}}$ to $\mathcal{U}_{\mathcal{A}}$.

The function h_c is well defined for all possible paths in $\Delta^{\mathcal{U}^c_{\mathcal{K}}}$. We prove it is a Σ-homomorphism from $\mathcal{U}^c_{\mathcal{K}}$ to $\mathcal{U}_{\mathcal{A}}$ by induction on the length of a path $\sigma \in \Delta^{\mathcal{U}^c_{\mathcal{K}}}$. Base of induction: $\mathbf{t}^{\mathcal{U}_{\mathcal{K}}}_{\Sigma}(c) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{A}}}_{\Sigma}(c)$ follows from $(c \mapsto c)$ is a state in $\mathsf{G}_c = (\mathsf{S}, \mathsf{D}, \mathsf{T})$. Step of induction. Let $\sigma \in \Delta^{\mathcal{U}^c_{\mathcal{K}}}$ be a path of length i and $\mathsf{tail}(\sigma) \rightsquigarrow_{\mathcal{K}} w_{[R]}$ for some role R. Moreover, assume $h_c(\sigma) = a$ and $h_c(\sigma w_{[R]}) = b$ for some constants $a, b \in \mathsf{Ind}(\mathcal{A})$. Then, it follows that there exist states

$$s = (\mathsf{tail}(\sigma) \mapsto a),\ s' = (a,\ \mathsf{tail}(\sigma) \rightsquigarrow w_{[R]}),\ \text{and } s'' = (w_{[R]} \mapsto b),$$

such that $(s, s'), (s', s'') \in \mathsf{T}$. By construction of T, from $(s', s'') \in \mathsf{T}$ it follows that $\mathbf{r}^{\mathcal{U}_{\mathcal{A}}}_{\Sigma}(a, b) \supseteq \mathbf{r}^{\mathcal{U}_{\mathcal{K}}}_{\Sigma}(\sigma, \sigma w_{[R]})$ and $\mathbf{t}^{\mathcal{U}_{\mathcal{A}}}_{\Sigma}(b) \supseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}}}_{\Sigma}(\sigma w_{[R]})$. By the inductive assumption, $\mathbf{t}^{\mathcal{U}_{\mathcal{A}}}_{\Sigma}(a) \supseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}}}_{\Sigma}(\sigma)$, hence, h_c is in fact, a Σ-homomorphism. Now, given (abox), a Σ-homomorphism from $\mathcal{U}_{\mathcal{K}}$ to $\mathcal{U}_{\mathcal{A}}$ can defined as the union of h_c for each $c \in \mathsf{Ind}(\mathcal{K})$. $\square$

The examples below illustrate the presented reduction.

Example 5.2.6. Consider $\mathcal{K}$, $\mathcal{A}$, Σ and the game $G^a_{\Sigma}(\mathcal{G}_{\mathcal{K}}, \mathcal{U}_{\mathcal{A}})$ from Example 5.2.4. In this case, there exists no homomorphism from $\mathcal{U}_{\mathcal{K}}$ to $\mathcal{U}_{\mathcal{A}}$, and as we have seen in Example 5.2.4, Spoiler has a winning strategy from $(a \mapsto a)$: in his first turn he should move to the state $(a,\ a \rightsquigarrow w_S)$. ■

Example 5.2.7. Assume $\Sigma = \{R'(\cdot,\cdot), Q'(\cdot,\cdot)\}$, $\mathcal{K} = \langle \mathcal{T}, \{\exists R(a), \exists S(a)\}\rangle$, where $\mathcal{T} = \{\exists S^- \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists Q, \exists Q^- \sqsubseteq \exists Q, R \sqsubseteq R', S \sqsubseteq R', Q \sqsubseteq Q'\}$ and $\mathcal{A} = \{R'(a,a), R'(a,b), Q'(b,b)\}$. Then $F_a = \{(b,\ w_S \rightsquigarrow w_R), (a,\ w_R \rightsquigarrow w_Q)\}$. In Figure 8 we depict the game graph G_a and a Σ-homomorphism h from $\mathcal{U}_{\mathcal{K}}$ to $\mathcal{U}_{\mathcal{A}}$.

Observe that in the game $G^a_{\Sigma}(\mathcal{G}_{\mathcal{K}}, \mathcal{U}_{\mathcal{A}})$ there is a way for Duplicator to play (infinitely) so that the game never goes out of his winning region W_D shown dotted in a). It is not difficult to see that such strategy of Duplicator can be used to define the homomorphism h, and vice versa. ■

Example 5.2.8. Assume $\Sigma = \{R'(\cdot,\cdot), S'(\cdot,\cdot), Q'(\cdot,\cdot)\}$, $\mathcal{K} = \langle \mathcal{T}, \{\exists R(a)\}\rangle$, where $\mathcal{T} = \{\exists R^- \sqsubseteq \exists S, \exists R^- \sqsubseteq \exists Q, R \sqsubseteq R', S \sqsubseteq S', Q \sqsubseteq Q'\}$, and $\mathcal{A} = \{R'(a,b), S'(b,c), R'(a,d), Q'(d,e)\}$. Then, the game graph G_a, the projection of $\mathcal{U}_{\mathcal{K}}$ over Σ and $\mathcal{U}_{\mathcal{A}}$ can be depicted as in Figure 9:

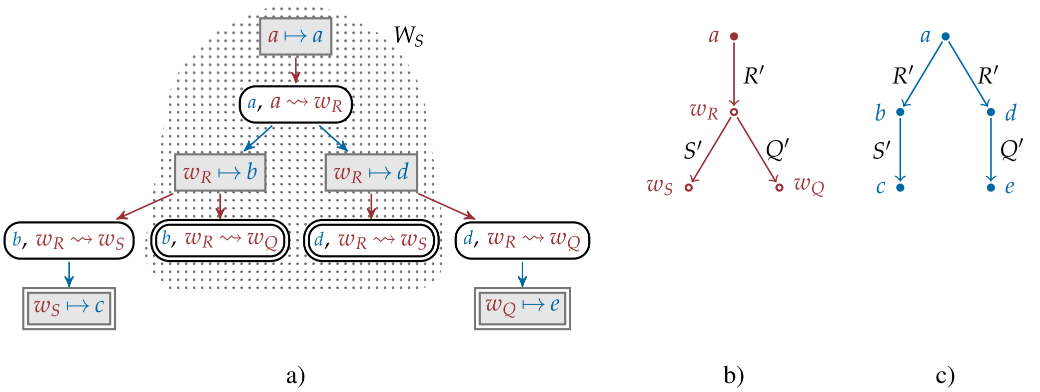


Figure 9: No homomorphism: a) the game graph G_a, b) the projection of $\mathcal{U}_\mathcal{K}$ over Σ, c) $\mathcal{U}_\mathcal{A}$

In this example, Spoiler has a winning strategy from $(a \mapsto a)$ (i.e., $(a \mapsto a)$ belongs to the winning region of Spoiler), and there is no homomorphism from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$. Note that here every possible play is finite. ■

Example 5.2.9. Assume $\Sigma = \{R'(\cdot,\cdot), S'(\cdot,\cdot)\}$, $\mathcal{K} = \langle \mathcal{T}, \{\exists Q(a)\}\rangle$, where $\mathcal{T} = \{\exists Q^- \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists S, R \sqsubseteq R', S \sqsubseteq S'\}$, and $\mathcal{A} = \{R'(a,c), S'(c,b), R'(b,c), R'(d,d)\}$. Then, the game graph G_a, the projection of $\mathcal{U}_\mathcal{K}$ over Σ and $\mathcal{U}_\mathcal{A}$ can be depicted as In Figure 10 we depict the game graph G_a and a Σ-homomorphism h from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$.

Observe that the projection of $\mathcal{U}_\mathcal{K}$ over Σ has a "break": $\mathbf{r}_\Sigma^{\mathcal{U}_\mathcal{K}}(a, aw_Q) = \emptyset$, so in principle aw_Q can be mapped to constants that are not connected to a in $\mathcal{U}_\mathcal{A}$. Hence, the state $(a,\ a \rightsquigarrow w_Q)$ in G_a has transitions to every state of the form $(w_Q \mapsto e)$ for $e \in \mathsf{Ind}(\mathcal{A})$. Note that we can do so because there is only a polynomial number of constants in $\mathcal{A}$.

In this example, Duplicator has two winning strategy from $(a \mapsto a)$ and there are two possible Σ-homomorphisms from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$. The depicted homomorphism h corresponds to the strategy that assigns to the state $(a,\ a \rightsquigarrow w_Q)$ the next state $(w_Q \mapsto a)$.

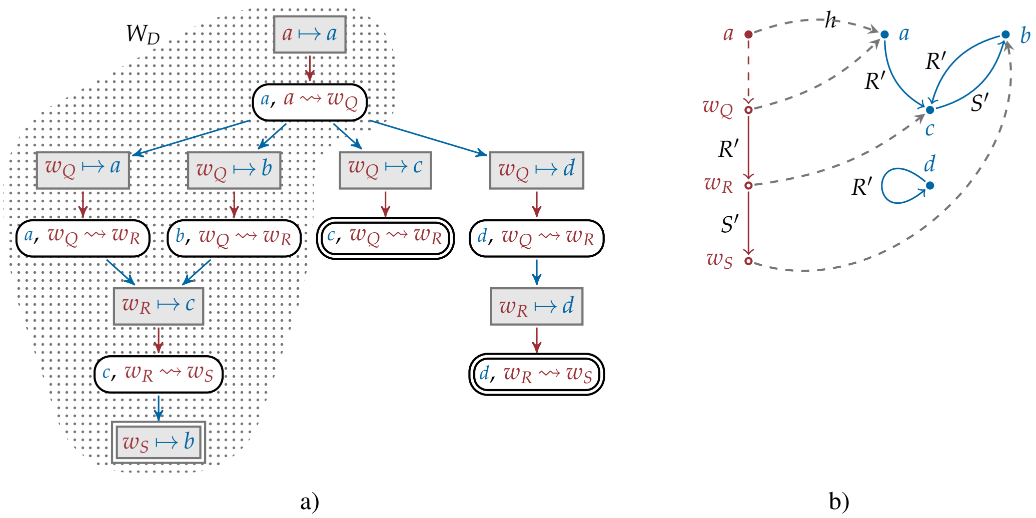


Figure 10: Example with "breaks" in $\mathcal{U}_\mathcal{K}$: a) G_a, b) Σ-homomorphism h from $\mathcal{U}_\mathcal{K}$ to $\mathcal{U}_\mathcal{A}$

■

It can be decided in polynomial time whether Duplicator has a winning strategy from a given state (in fact, in linear time [33]), therefore condition (win) can be checked in polynomial time. Finally, combining Lemma 5.2.5, Lemma 5.1.2, and Proposition A.1.1 one obtains:

Lemma 5.2.10. *The membership problem for simple universal solutions is in* PTIME.

We conclude this section with the exact complexity for the membership and non-emptiness problems for simple universal solutions. The following is a consequence of Lemmas 5.2.1, 5.2.3, and 5.2.10.

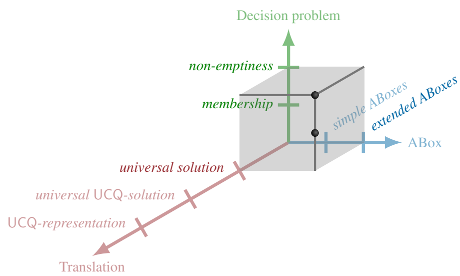
Theorem 5.2.11. *The non-emptiness problem for simple universal solutions is* PTIME-*complete. Moreover, there is an effective algorithm to compute a universal solution in polynomial time (if such a solution exists).*

One case see that the reduction in Lemma 5.2.1 can be easily adapted to the case of the membership problem, so we obtain PTIME-completeness also for this problem.

Theorem 5.2.12. *The membership problem for simple universal solutions is* PTIME-*complete.*

5.3 EXTENDED UNIVERSAL SOLUTIONS

In this section we study the membership and the non-emptiness problems for universal solutions when extended ABoxes are allowed in the target. In Section 5.3.1, we show that the former problem is NP-complete, while in Section 5.3.2, we provide the non-matching bounds for the latter problem, namely a PSPACE lower bound and an EXPTIME upper bound.



5.3.1 *The membership problem*

Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ over Σ_1, and an extended ABox $\mathcal{A}_2$ over Σ_2. In this setting, Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$ can be still checked in PTIME using the technique of reachability games presented in Section 5.2, however, the opposite direction cannot be checked efficiently due to nulls in $\mathcal{A}_2$. In fact, it can be shown by reduction from the graph 3-colorability problem that the membership problem for universal solutions with null values is NP-hard.

Lemma 5.3.1. *The membership problem for extended universal solutions is* NP-*hard in data complexity.*

Proof. The proof is by reduction from 3-colorability of undirected graphs known to be NP-hard. Consider an undirected graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$, and fix signatures $\Sigma_1 = \{E(\cdot,\cdot)\}$ and $\Sigma_2 = \{E'(\cdot,\cdot)\}$. Further, let $r, g, b \in N_a$, $\mathsf{V} \subseteq N_l$ and

$$\begin{aligned}
\mathcal{A}_1 &= \{E(r,g), E(g,r), E(r,b), E(b,r), E(g,b), E(b,g)\},\\
\mathcal{T}_1 &= \{\},\\
\mathcal{T}_{12} &= \{E \sqsubseteq E'\},\\
\mathcal{A}_2 &= \{E'(r,g), E'(g,r), E'(r,b), E'(b,r), E'(g,b), E'(b,g)\}\\
&\quad \cup \{E'(x,y), E'(y,x) \mid (x,y) \in \mathsf{E}\}.
\end{aligned}$$

Note that the nodes in G become labeled nulls in $\mathcal{A}_2$. We show that G is 3-colorable if and only if $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$.

($\Rightarrow$) Suppose G is 3-colorable. Then it follows that there exists a function h that assigns to each vertex from V one of the colors $\{r, g, b\}$ such that if $(x,y) \in \mathsf{E}$, then $h(x) \neq h(y)$, hence h is a homomorphism from G to the undirected graph $(\{r,g,b\}, \{(r,g),(g,b),(b,r)\})$.

We prove that $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Obviously, $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$. Thus, it remains to verify that $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. First, it is easy to see that $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically embeddable into $\mathcal{U}_{\mathcal{A}_2}$. Second, h is also a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, thus $\mathcal{U}_{\mathcal{A}_2}$ is homomorphically embeddable into $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$.

($\Leftarrow$) Suppose now $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Then by Lemma 5.1.2 it follows that $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Let h be a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Notice that $\Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} = \mathsf{Ind}(\mathcal{A}_1)$, hence h assigns to each labeled null $x \in \Delta^{\mathcal{U}_{\mathcal{A}_2}}$ some constant $a \in \mathsf{Ind}(\mathcal{A}_1)$, and it is easy to see that h is an assignment for the vertices in V that is a 3-coloring of G. □

To decide in NP whether there exists a homomorphism h from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, we can use the fact that the image $W \subseteq \Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ of such a function h on $\Delta^{\mathcal{U}_{\mathcal{A}_2}}$ is of polynomial size. Therefore, for each constant and null in $\mathcal{A}_2$, one needs to guess its homomorphic image, and then check whether the resulting function is a homomorphism. Thus, we obtain an NP-upper bound for the membership problem for universal solutions with extended ABoxes.

Lemma 5.3.2. *The membership problem for extended universal solutions is in* NP.

Proof. Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a source KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, and an extended target ABox $\mathcal{A}_2$, and the question to answer is whether $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. It is sufficient to show that condition (hom) of Lemma 5.1.2 can be checked in NP. The existence of a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}_{\mathcal{A}_2}$ can be decided in PTIME using the technique of reachability games presented in Section 5.2 (note that for homomorphisms in this direction, there is no distinction made between the constants and the labeled nulls in $\mathcal{A}_2$). In the rest of this proof, we show how to check the existence of a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ in NP in the size of $\mathcal{K}_1$, $\mathcal{M}$ and $\mathcal{A}_2$.

First, if there exists a homomorphism h from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, then there exists a witness W with a number of elements bounded by the size of $\mathcal{A}_2$ such that $W \subseteq \Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ and h is a function from $\Delta^{\mathcal{U}_{\mathcal{A}_2}}$ to W: take $W = h(\Delta^{\mathcal{U}_{\mathcal{A}_2}})$.

Second, we show that there exists a witness W such that $W \subseteq \Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ and every $x \in W$ is a path of length smaller or equal $2m$, where for $x = aw_{[S_1]} \cdots w_{[S_k]}$ the length of x is $k+1$, and m is the size of $\mathcal{T}_1 \cup \mathcal{T}_{12} \cup \mathcal{A}_2$. To this purpose, let h be a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ and $W = h(\Delta^{\mathcal{U}_{\mathcal{A}_2}})$. For $x, y \in W$, we say that x is connected to y in $\mathcal{U}^W_{\mathcal{K}}$, if there exists $n \geq 0$ and a path $(x_1, x_2, \ldots, x_n, x_{n+1})$ such that $x_i \in W$, $x_1 = x$, $x_{n+1} = y$, and $(x_i, x_{i+1}) \in R_i^{\mathcal{U}^W_{\mathcal{K}}}$ for some role R_i, for $i \in \{1, \ldots, n\}$. Assume that $x \in W$ and the length of x is more than $2m$. Then, since $W = h(\Delta^{\mathcal{U}_{\mathcal{A}_2}})$, we have that x is not connected to any element of $\mathsf{Ind}(\mathcal{A}_1)$ in $\mathcal{U}^W_{\mathcal{K}}$. Let C be the maximal connected subset of W with $x \in C$, i.e., for each $y \in C$, *(i)* y is connected to y' in $\mathcal{U}^W_{\mathcal{K}}$, for each $y' \in C$, and *(ii)* y is not connected to any $z \in W \setminus C$. Note that $C \cap \mathsf{Ind}(\mathcal{A}_1) = \emptyset$. Let y be the path (in the sense of $\mathsf{path}(\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle)$) of minimal length in C, it exists and is unique since $C \subseteq \Delta^{\mathcal{U}^W_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ and there are no constants in C. Then for each $y' \in C$, $y' = y \cdot w_{[R_1]} \ldots w_{[R_n]}$ for some roles $R_1, \ldots, R_n$. Further assume $\mathsf{tail}(y) = w_{[R]}$, and let z be a path of the minimal length in $\Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ with $\mathsf{tail}(z) = w_{[R]}$. Then the length of z is bounded by the size of $\mathcal{T}_1 \cup \mathcal{T}_{12}$ and the length of each $z \cdot w_{[R_1]} \ldots w_{[R_n]}$ for some $y \cdot w_{[R_1]} \ldots w_{[R_n]} \in C$, is bounded by the size of $\mathcal{T}_1 \cup \mathcal{T}_{12} \cup \mathcal{A}_2$. Now, define a new function $h' : \Delta^{\mathcal{U}_{\mathcal{A}_2}} \to \Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ such that $h'(x) = h(x)$ if $h(x) \notin C$, and $h'(x) = z \cdot w_{[R_1]} \ldots w_{[R_n]}$ if $h(x) = y \cdot w_{[R_1]} \ldots w_{[R_n]}$. It is easy to see that h' is a homomorphism from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. Now, we can take $W = h'(\Delta^{\mathcal{U}_{\mathcal{A}_2}})$, and repeat the above construction until the claim is satisfied.

Finally, to verify in NP whether a homomorphism h from $\mathcal{U}_{\mathcal{A}_2}$ to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ exists, it is sufficient to guess W as above and then to check whether $\mathcal{U}_{\mathcal{A}_2}$ can be homomorphically mapped to $\mathcal{U}^W_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. □

Thus, we obtain the exact complexity bound for the membership problem with extended ABoxes.

Theorem 5.3.3. *The membership problem for extended universal solutions is* NP*-complete.*

5.3.2 *The non-emptiness problem*

Consider now the non-emptiness problem for universal solutions with null values, that is, when extended ABoxes are allowed in universal solutions. This problem turns out to be harder than the membership problem as now candidate solutions that can be of exponential size are not part of the input. In fact, we show by reduction from the validity problem for quantified Boolean formulas that checking the existence of a universal solution is PSPACE-hard. Then we provide an EXPTIME algorithm based on two-way alternating tree automata (2ATA), known to be EXPTIME-complete and defined in Section A.1.1. 2ATA have been previously successfully applied not only in the domain of μ-calculus and temporal logics [46, 21, 45], but also in the DLs domain to reason and perform query answering in expressive DLs [29, 31, 32].

THE LOWER BOUND

Lemma 5.3.4. *The non-emptiness problem for extended universal solutions is* PSPACE*-hard.*

Proof. The proof is by reduction from the validity problem for quantified Boolean formulas, known to be PSPACE-complete. Suppose we are given a QBF

$$\phi = \mathsf{Q}_1 X_1 \ldots \mathsf{Q}_n X_n \bigwedge_{j=1}^{m} C_j$$

where $\mathsf{Q}_i \in \{\forall, \exists\}$ and C_j, $1 \leq j \leq m$, are clauses over the variables X_i, $1 \leq i \leq n$.

Let $\Sigma_1 = \{A, Y_i^k, X_i^k, S_l, T_l, Q_i^k, P_i^k, R_j, R_j^l \mid 1 \leq j \leq m, 1 \leq i \leq n, 0 \leq l \leq n, k \in \{0,1\}\}$ where A, Y_i^k, X_i^k are concept names and the rest are role names. Let $\mathcal{T}_1$ be the following TBox over Σ_1 for $1 \leq j \leq m$, $1 \leq i \leq n$ and $k \in \{0,1\}$:

$$\begin{array}{llll}
A \sqsubseteq \exists S_0^- & \exists S_{i-1}^- \sqsubseteq \exists Q_i^k \ \text{ if } \mathsf{Q}_i = \forall & Q_i^k \sqsubseteq S_i & \exists (Q_i^k)^- \sqsubseteq Y_i^k \\
 & \exists S_{i-1}^- \sqsubseteq \exists S_i \ \text{ if } \mathsf{Q}_i = \exists & \exists S_n^- \sqsubseteq \exists R_j & \exists R_j^- \sqsubseteq \exists R_j
\end{array}$$

$$\begin{array}{llll}
A \sqsubseteq \exists T_0^- & \exists T_{i-1}^- \sqsubseteq \exists P_i^k & X_i^0 \sqsubseteq \exists R_j^i \ \text{ if } \neg X_i \in C_j & \exists (R_j^i)^- \sqsubseteq \exists R_j^{i-1} \\
P_i^k \sqsubseteq T_i & \exists (P_i^k)^- \sqsubseteq X_i^k & X_i^1 \sqsubseteq \exists R_j^i \ \text{ if } X_i \in C_j &
\end{array}$$

and $\mathcal{A}_1 = \{A(a)\}$.

Further, let $\Sigma_2 = \{Z_i^0, Z_i^1, S', R_j'\}$ where Z_i^0, Z_i^1 are concept names and S', R_j' are role names, $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, and $\mathcal{T}_{12}$ the following set of inclusions:

$$\begin{array}{llll}
S_i \sqsubseteq S' & Y_i^k \sqsubseteq Z_i^k & R_j \sqsubseteq R_j' & R_j^i \sqsubseteq R_j' \\
T_i \sqsubseteq S' & X_i^k \sqsubseteq Z_i^k & T_i \sqsubseteq R_j'^- & R_j^0 \sqsubseteq R_j'^-
\end{array}$$

Then, $\models \phi$ if and only if $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically embeddable into a finite subset of itself, i.e., if and only if a universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$ exists. The rest of the proof follows the line of the proof of Theorem 11 in [77].

($\Rightarrow$) Suppose $\models \phi$. We show that the canonical model $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically embeddable into a finite subset of itself. More precisely, let us denote with $\mathcal{T}_1^{inf}$ the subset of $\mathcal{T}_1$ consisting of the first 9 axioms, and $\mathcal{T}_1^{fin}$ the subset of $\mathcal{T}_1$ consisting of the last 9 axioms. Then $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle} = \mathcal{U}_{\langle \mathcal{T}_1^{inf} \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle} \cup \mathcal{U}_{\langle \mathcal{T}_1^{fin} \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, and we construct a Σ_2-homomorphism $h : \Delta^{\mathcal{U}_{\langle \mathcal{T}_1^{inf} \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}} \to \Delta^{\mathcal{U}_{\langle \mathcal{T}_1^{fin} \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$. In the following we use $\mathcal{I}$ to denote $\mathcal{U}_{\langle \mathcal{T}_1^{inf} \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, and $\mathcal{F}$ to denote $\mathcal{U}_{\langle \mathcal{T}_1^{fin} \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$.

We begin by setting $h(a^{\mathcal{I}}) = a^{\mathcal{F}}$. Then we define h in such a way that, for each path π in $\mathcal{I}$ of length $i+1 \leq n$, $h(\pi)$ is a path $a^{\mathcal{F}} w_1 \ldots w_i$ of length $i+1$ in $\mathcal{F}$ and it defines an assignment $\mathfrak{a}_{h(\pi)}$ to the variables $X_1, \ldots, X_i$ by taking, for all $1 \leq i' \leq i$,

$$\begin{array}{l}
\mathfrak{a}_{h(\pi)}(X_{i'}) = \top \Leftrightarrow a^{\mathcal{F}} \cdot w_1 \cdot \ldots \cdot w_{i'} \in (X_{i'}^1)^{\mathcal{F}} \\
\mathfrak{a}_{h(\pi)}(X_{i'}) = \bot \Leftrightarrow a^{\mathcal{F}} \cdot w_1 \cdot \ldots \cdot w_{i'} \in (X_{i'}^0)^{\mathcal{F}}.
\end{array}$$

Such assignments $\mathfrak{a}_{h(\pi)}$ will satisfy the following:

($\mathfrak{a}$) the QBF obtained from ϕ by removing $\mathsf{Q}_1 X_1 \ldots \mathsf{Q}_i X_i$ from its prefix is true under $\mathfrak{a}_{h(\pi)}$.

For the paths of length 0 the Σ_2-homomorphism h has been defined and ($\mathfrak{a}$) trivially holds. Suppose that we have defined h for all paths in $\mathcal{I}$ of length $i+1 \leq n$. We extend

h to all paths of length $i+2$ in $\mathcal{I}$ such that $(\mathfrak{a})$ holds. Let π be a path of length $i+1$. In $\mathcal{F}$ we have

$$\mathsf{tail}(h(\pi)) \rightsquigarrow_{\langle \mathcal{T}_1^{fin} \cup \mathcal{T}_{12}, \mathcal{A}_2 \rangle} w^{\mathcal{F}}_{[P_i^k]}, \quad \text{and} \quad h(\pi) \cdot w^{\mathcal{F}}_{[P_i^k]} \in (X_i^k)^{\mathcal{F}}, \text{ for } k = 0, 1.$$

If $\mathsf{Q}_i = \forall$ then in $\mathcal{I}$ we have

$$\mathsf{tail}(\pi) \rightsquigarrow_{\langle \mathcal{T}_1^{inf} \cup \mathcal{T}_{12}, \mathcal{A}_2 \rangle} w^{\mathcal{I}}_{[Q_i^k]}, \quad \text{and} \quad \pi \cdot w^{\mathcal{I}}_{[Q_i^k]} \in (X_i^k)^{\mathcal{I}}, \text{ for } k = 0, 1.$$

Thus, we set $h(\pi \cdot w^{\mathcal{I}}_{[Q_i^k]}) = h(\pi) \cdot w^{\mathcal{F}}_{[P_i^k]}$, for $k = 0, 1$. Clearly, $(\mathfrak{a})$ holds. Otherwise, $\mathsf{Q}_i = \exists$ and in $\mathcal{I}$ we have

$$\mathsf{tail}(\pi) \rightsquigarrow_{\langle \mathcal{T}_1^{inf} \cup \mathcal{T}_{12}, \mathcal{A}_2 \rangle} w^{\mathcal{I}}_{[S_i]}.$$

We know that $\models \phi$ and so, by, $(\mathfrak{a})$, the QBF obtained from π by removing $\mathsf{Q}_1 X_1 \ldots \mathsf{Q}_i X_i$ is true under either $\mathfrak{a}_{h(\pi)} \cup \{X_i = \top\}$ or $\mathfrak{a}_{h(\pi)} \cup \{X_i = \bot\}$. We set $h(\pi \cdot w^{\mathcal{I}}_{[S_i]}) = h(\pi) \cdot w^{\mathcal{F}}_{[P_i^k]}$ with $k = 1$ in the former case, and $k = 0$ in the latter case. Either way, $(\mathfrak{a})$ holds.

Consider now in $\mathcal{I}$ a path π of length $n+1$ from $a^{\mathcal{I}}$ to $w_n^{\mathcal{I}}$. By construction, we have

$$h(\pi) = a^{\mathcal{F}} \cdot w^{\mathcal{F}}_{[P_1^{k_1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[P_n^{k_n}]}.$$

Next, on the one hand, the path π in $\mathcal{I}$ has m infinite extensions of the form $\pi \cdot w^{\mathcal{I}}_{[R_j]} \cdot w^{\mathcal{I}}_{[R_j]} \ldots$, for $1 \leq j \leq m$. On the other hand, as $\models \phi$, by $(\mathfrak{a})$, for each clause C_j, there is some $1 \leq i' \leq n$ such that $h(\pi)$ contains $w^{\mathcal{F}}_{[P_{i'}^1]}$ if $X_{i'} \in C_j$, or $w^{\mathcal{F}}_{[P_{i'}^0]}$ if $\neg X_{i'} \in C_j$. We set for each $1 \leq l \leq n - i'$,

$$h(\pi \cdot \underbrace{w^{\mathcal{I}}_{[R_j]} \cdot \ldots \cdot w^{\mathcal{I}}_{[R_j]}}_{l \text{ times}}) = a^{\mathcal{F}} \cdot w^{\mathcal{F}}_{[P_1^{k_1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[P_{n-l}^{k_{n-l}}]},$$

for each $n+1 \geq l > n - i'$,

$$h(\pi \cdot \underbrace{w^{\mathcal{I}}_{[R_j]} \cdot \ldots \cdot w^{\mathcal{I}}_{[R_j]}}_{l \text{ times}}) = a^{\mathcal{F}} \cdot w^{\mathcal{F}}_{[P_1^{k_1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[P_{i'}^{k_{i'}}]} \cdot w^{\mathcal{F}}_{[R_j^{i'}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[R_j^{n-l+1}]},$$

and for each $l > n+1$

$$h(\pi \cdot \underbrace{w^{\mathcal{I}}_{[R_j]} \cdot \ldots \cdot w^{\mathcal{I}}_{[R_j]}}_{l \text{ times}}) = a^{\mathcal{F}} \cdot w^{\mathcal{F}}_{[P_1^{k_1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[P_{i'}^{k_{i'}}]} \cdot w^{\mathcal{F}}_{[R_j^{i'}]} \cdot w^{\mathcal{F}}_{[R_j^{i'-1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[R_j^{i^\star}]},$$

where $i^\star = (n - l + 1) \bmod 2$. It is immediate to verify that h is a Σ_2-homomorphism from $\mathcal{I}$ to $\mathcal{F}$. Since $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, by Lemma 5.1.6 we obtain that a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$ exists.

($\Leftarrow$) Let h be a Σ_2-homomorphism from $\mathcal{I}$ to $\mathcal{F}$. We show that $\models \phi$.

Let π be a path of length $n+1$, $\pi = a^{\mathcal{I}} \cdot w_1 \cdot \ldots \cdot w_n$, in $\mathcal{I}$. Then $(a^{\mathcal{I}}, \pi_1), (\pi_i, \pi_{i+1}) \in S'^{\mathcal{I}}$, where $\pi_i = a^{\mathcal{I}} \cdot w_1 \cdot \ldots \cdot w_i$, for $1 \leq i \leq n-1$. Furthermore, let $Z_1^{k_1}, Z_2^{k_2}, \ldots, Z_n^{k_n}$ be the concepts containing subpaths of $h(\pi_i)$. We show that for every $1 \leq j \leq m$, the clause C_j contains at least one of the literals

$$\{X_i \mid k_i = 1, 1 \leq i \leq n\} \cup \{\neg X_i \mid k_i = 0, 1 \leq i \leq n\}.$$

Validity of ϕ will follow.

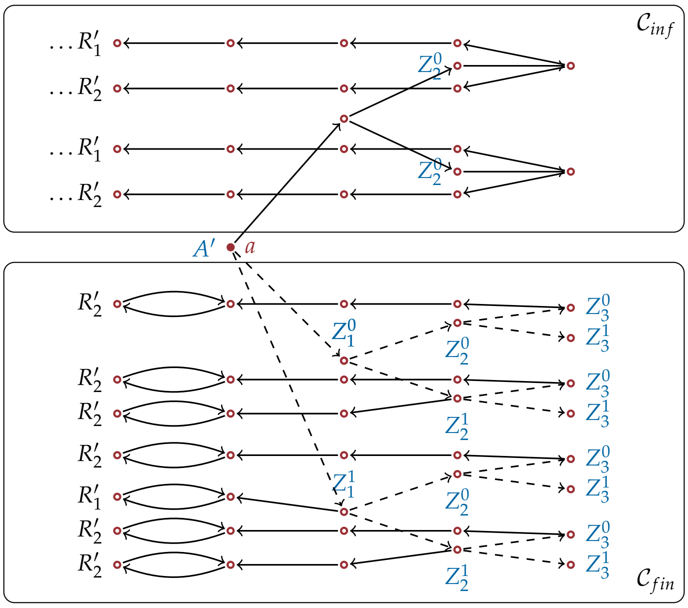
Consider a path of the form $\pi \cdot \underbrace{w^{\mathcal{I}}_{[R_j]} \cdot \ldots \cdot w^{\mathcal{I}}_{[R_j]}}_{n+1 \text{ times}}$ in $\mathcal{I}$. Then its h-image in $\mathcal{F}$ must be of the form

$$a^{\mathcal{F}} \cdot w^{\mathcal{F}}_{[P_1^{k_1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[P_i^{k_i}]} \cdot w^{\mathcal{F}}_{[R_j^i]} \cdot w^{\mathcal{F}}_{[R_j^{i-1}]} \cdot \ldots \cdot w^{\mathcal{F}}_{[R_j^{i'}]}$$

for some $1 \leq i \leq n$, $i' = 0$ or $i' = 1$, and $k_i = 0$ or $k_i = 1$. If $k_i = 0$, then C_j must contain $\neg X_i$, otherwise X_i. □

The following example illustrates the reduction.

Example 5.3.5. For $\phi = \exists X_1 \forall X_2 \exists X_3 (X_1 \wedge (X_2 \vee \neg X_3))$, the projection of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ over Σ_2 can be depicted as follows:



where each edge $\longrightarrow$ is labeled with S', each edge $\dashrightarrow$ is labeled with $S', R_j'^-$ for $1 \leq j \leq m$, and the labels of edges $\longrightarrow$ are shown to the left of each infinite and finite path. The labels of the nodes (if any) are shown next to each node.

One can see that ϕ is valid. Let $\mathcal{C}_{inf}$ and $\mathcal{C}_{fin}$ be the parts of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ generated using the first 7 axiom templates and the last 7 axiom templates of $\mathcal{T}_1$ respectively. Note that $\mathcal{C}_{inf}$ is infinite, while $\mathcal{C}_{fin}$ is finite. Then $\mathcal{C}_{inf}$ is Σ_2-homomorphically embeddable into $\mathcal{C}_{fin}$ (hence, $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically embeddable into $\mathcal{C}_{fin}$), and the ABox obtained from $\mathcal{C}_{fin}$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

THE UPPER BOUND Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a source KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, and the question to decide is whether there exists a(n extended) universal solution for $\mathcal{K}$ under $\mathcal{M}$. We show how to check condition (core) of Lemma 5.1.6, that is, whether there exists a finite subset D of $\Delta^{\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}}$ and a Σ_2-homomorphism from $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ to $\mathcal{U}^D_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. To simplify the presentation, in the rest of this section we tackle a more general problem: given two (non-extended) KBs $\mathcal{K}_1$ and $\mathcal{K}_2$ with universal models $\mathcal{U}_1$ and $\mathcal{U}_2$, and a signature Σ, decide whether there exists a finite subset D of $\Delta^{\mathcal{U}_2}$ and a Σ-homomorphism from $\mathcal{U}_1$ to $\mathcal{U}_2^D$.

As in the case of the membership problem for simple universal solutions in Section 5.2, for such a homomorphism to exist, first, it should be the case that an analog of condition (abox) holds. Second, for $c \in \mathsf{Ind}(\mathcal{K}_1)$, to check whether the tree $\mathcal{U}_1^c$ can be Σ-homomorphically embedded into $\mathcal{U}_2^D$ for some finite $D \subseteq \Delta^{\mathcal{U}_2}$, we adopt *two-way alternating automata on infinite trees (2ATA)*, which are a generalization of nondeterministic automata on infinite trees [102] well suited for handling inverse roles in *DL-Lite*$_\mathcal{R}$. It is known that the non-emptiness problem for such automata is in EXPTIME, thus, we obtain the required upper bound for our problem. More precisely, we show for each constant $c \in \mathsf{Ind}(\mathcal{K}_1)$ how to construct automata $\mathbb{A}_c$ (with Büchi acceptance condition) such that its language is non-empty if and only if there exists a Σ-homomorphism from $\mathcal{U}_1^c$ to $\mathcal{U}_2^D$ for some finite $D \subseteq \Delta^{\mathcal{U}_2}$, and it accepts a tree if it corresponds to such a $\mathcal{U}_2^D$. Then to verify that there exists a Σ-homomorphism from $\mathcal{U}_1$ to $\mathcal{U}_2^D$ for some finite $D \subseteq \Delta^{\mathcal{U}_2}$, we solve the non-emptiness problem of $\mathbb{A}_c$ for each $c \in \mathsf{Ind}(\mathcal{K}_1)$. If the language accepted by some $\mathbb{A}_c$ is empty, then there is no such homomorphism, otherwise we can compute $\mathcal{U}_2^D$ as the union of the trees accepted by $\mathbb{A}_c$. Below we show how to construct the automaton $\mathbb{A}_c$.

AUTOMATON $\mathbb{A}_c$ Let $\mathsf{Ind}(\mathcal{K}_2) = \{a_1, \ldots, a_{n_a}\}$, $\mathsf{Wit}(\mathcal{K}_2) = \{w_1, \ldots, w_{n_w}\}$, and $n = \max(n_a, n_w)$. Moreover, denote by $\mathcal{G}_1$ and $\mathcal{G}_2$ the generating models of $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively.

We define *automaton* $\mathbb{A}_c$ as the tuple $\langle \Gamma, Q, \delta, q_0, F \rangle$, where the *alphabet* Γ is the set

$$\Gamma = \{R, S\} \cup \{\overline{a}_i \mid 1 \leq i \leq n_a\} \cup \{\overline{w}_i \mid 1 \leq i \leq n_w\}.$$

Hence, $\mathbb{A}_c$ accepts n-ary trees where each node either corresponds to a constant of $\mathcal{K}_2$, then it should be labeled with the symbol $\overline{a}_i$, or corresponds to a witness of $\mathcal{K}_2$, labeled with the symbol $\overline{w}_i$, or is the root of the tree, labeled with R, or is a node outside the finite part, labeled with S (S stands for "stop"). The set Q of *states* is partitioned into three sets:

$$Q = \{q_0\} \cup Q_f \cup Q_h,$$

where Q_f is the set of states responsible for labeling the tree as an appropriate finite substructure of $\mathcal{U}_2$, and Q_h is the set of states responsible for checking the homomorphism from $\mathcal{U}_1$ into a finite substructure of $\mathcal{U}_2$. Thus,

$$Q_f = \{q_f\} \cup \{\alpha_i \mid 1 \leq i \leq n_a\} \cup \{\omega_i \mid 1 \leq i \leq n_w\},$$

where the states α_i are responsible for labeling the tree with the constants of $\mathcal{K}_2$ and ω_i are responsible for labeling the tree with the witnesses of $\mathcal{K}_2$. We define the transition function for these states and q_0 as follows:

$$\delta(q_0, L) = (0, q_f) \wedge (0, q_h), \tag{2}$$

$$\delta(q_f, L) = \begin{cases} \bigwedge_{i=1}^{n_a} (i, \alpha_i), & \text{if } L = R \\ \bot, & \text{otherwise,} \end{cases} \tag{3}$$

$$\text{for } 1 \leq i \leq n_a, \quad \delta(\alpha_i, L) = \begin{cases} \bigwedge\limits_{\substack{1 \leq j \leq n_w, \\ a_i \rightsquigarrow_{\mathcal{K}_2} w_j}} (j, \omega_j), & \text{if } L = \overline{a}_i \\ \bot, & \text{otherwise} \end{cases} \tag{4}$$

$$\text{for } 1 \leq i \leq n_w, \quad \delta(\omega_i, L) = \begin{cases} \bigwedge\limits_{\substack{1 \leq j \leq n_w, \\ w_i \rightsquigarrow_{\mathcal{K}_2} w_j}} (j, \omega_j), & \text{if } L = \overline{w}_i \\ \top, & \text{if } L = S \\ \bot, & \text{if } L \in \Gamma \setminus \{\overline{w}_i, S\} \end{cases} \tag{5}$$

where q_h is a state from Q_h we are going to define below. For now observe that due to the transitions above the trees accepted by $\mathbb{A}_c$ will have the symbol R in the root and the symbol $\overline{a}_i$ in the i-th successor of the root. Then each of the i-th successors above will have its j-th successor marked with $\overline{w}_j$ whenever $a_i \rightsquigarrow_{\mathcal{K}_2} w_j$. Further each of the j-th successors above will have its i-th successor marked with $\overline{w}_i$ whenever $w_j \rightsquigarrow_{\mathcal{K}_2} w_i$, and so on. Note that at some step the node in the tree marked with $\overline{w}_j$ can have its i-th successor marked with S (instead of $\overline{w}_i$), when $w_j \rightsquigarrow_{\mathcal{K}_2} w_i$. This should mean that this i-th successor is not inside the finite substructure of $\mathcal{U}_2$ to which the homomorphism will be mapped, and the automata must stop going down the tree. Note that it is not yet guaranteed that S appears instead of $\overline{w}_i$ at some point, however, if it is not the case, we may have an infinite path in the tree, in which the automata will pass through the states $\omega_{i_1}, \omega_{i_2}, \ldots$ and that would be a contradiction to our acceptance condition presented below.

To define Q_h we need some additional notation. Let $\mathsf{Wit}(\mathcal{K}_1) = \{u_1, \ldots, u_m\}$, and assume that u_0 is used to denote c. Let $t, t' \in \{u_0, \ldots, u_m\}$. In the definition of transitions of the automata we will use a pair of relations $\xrightarrow[\Sigma]{}$ and $\underset{\Sigma}{\dashrightarrow}$ between t and t' defined, respectively, as $t \xrightarrow[\Sigma]{} t'$ if and only if $t \rightsquigarrow_{\mathcal{K}_1} t'$ and $\mathbf{r}_{\Sigma}^{\mathcal{G}_1}(t, t') \neq \emptyset$, and $t \underset{\Sigma}{\dashrightarrow} t'$ if and only if $t \rightsquigarrow_{\mathcal{K}_1} t'$ and $\mathbf{r}_{\Sigma}^{\mathcal{G}_1}(t, t') = \emptyset$. One needs to distinguish these two relations: suppose an element $cu_{l_1} \cdots u_{l_k}$ of $\Delta^{\mathcal{U}_1}$ is homomorphically mapped to the element $a_{i_1} w_{i_2} \cdots w_{i_r}$ of $\Delta^{\mathcal{U}_2}$ and $u_{l_k} \rightsquigarrow_{\mathcal{K}_1} u_{l_{k+1}}$. If $u_{l_k} \xrightarrow[\Sigma]{} u_{l_{k+1}}$ then the element $cu_{l_1} \cdots u_{l_k} u_{l_{k+1}}$ of $\Delta^{\mathcal{U}_1}$ has to be mapped to an immediate successor or predecessor of the image of $cu_{l_1} \cdots u_{l_k}$ in $\mathcal{U}_2$. If, however, $u_{l_k} \underset{\Sigma}{\dashrightarrow} u_{l_{k+1}}$ then $cu_{l_1} \ldots u_{l_k} u_{l_{k+1}}$ can be mapped to any element of $\mathcal{U}_2$. Thus, the transitions of the automata on Q_h should be defined to reflect this.

Further, let $s, s' \in \{a_1, \ldots, a_{n_a}, w_1, \ldots, w_{n_w}\}$. We define a function $\rho_{s,s'}^{t,t'}$ between the pairs t, t' and s, s' such that $\rho_{s,s'}^{t,t'} = \top$ if $\mathbf{r}_{\Sigma}^{\mathcal{G}_1}(t, t') \subseteq \mathbf{r}_{\Sigma}^{\mathcal{G}_2}(s, s')$, and $\rho_{s,s'}^{t,t'} = \bot$ otherwise. Additionally for t and s, define a function τ_s^t such that $\tau_s^t = \top$ if $\mathbf{t}_{\Sigma}^{\mathcal{G}_1}(t) \subseteq \mathbf{t}_{\Sigma}^{\mathcal{G}_2}(s)$, and $\tau_s^t = \bot$ otherwise. Clearly, $\rho_{w_{i_r}, w_{i_{r+1}}}^{u_{l_k}, u_{l_{k+1}}} = \top$ and $\tau_{w_{i_{r+1}}}^{u_{l_{k+1}}} = \top$ in the example above guarantees the edge $(cu_{l_1} \cdots u_{l_k},\ cu_{l_1} \cdots u_{l_k} u_{l_{k+1}})$ of $\mathcal{U}_1$ can be mapped to the edge $(a_{i_1} w_{i_2} \cdots w_{i_r},\ a_{i_1} w_{i_2} \cdots w_{i_r} w_{i_{r+1}})$ of $\mathcal{U}_2$. Finally, we need a function $\eta_w^{u,u'}$ for $u, u' \in \{u_1, \ldots, u_m\}$ and $w \in \{w_1, \ldots, w_{n_w}\}$ such that $\eta_w^{u,u'} = \top$ if $\{R^- \mid R \in \mathbf{r}_{\Sigma}^{\mathcal{G}_1}(u, u')\} \subseteq \mathbf{r}_{\Sigma}^{\mathcal{G}_2}(s, w)$ for some $s \rightsquigarrow_{\mathcal{K}_2} w$, and $\eta_w^{u,u'} = \bot$ otherwise. In the example above $\eta_{w_{i_r}}^{u_{l_k}, u_{l_{k+1}}} = \top$ means that the edge $(cu_{l_1} \cdots u_{l_k},\ cu_{l_1} \cdots u_{l_k} u_{l_{k+1}})$ can be "inversely" mapped to the edge $(a_{i_1} w_{i_2} \cdots w_{i_r},\ a_{i_1} w_{i_2} \cdots w_{i_{r-1}})$, otherwise it cannot. Thus,

$$Q_h = \{q_h\} \cup \{\gamma_l,\ \chi_l \mid 0 \leq l \leq m\} \cup \{\kappa_l^i \mid 1 \leq l \leq m, 1 \leq i \leq n_a\}.$$

For each witness $u_l \in \mathsf{Wit}(\mathcal{K}_1)$ there are two states: γ_l is responsible for finding a homomorphic image, and χ_l is the "expecting state"; moreover for each witness $u_l \in \mathsf{Wit}(\mathcal{K}_1)$ and constant $a_i \in \mathsf{Ind}(\mathcal{K}_2)$ there is a state κ_l^i responsible for transition from some constant to a_i via the root and then mapping u_l to a_i. Intuitively, having a *run* of the automata on a tree $(\{1,\ldots,n\}^*, V)$ with a prefix

$$(\epsilon, q_0), (\epsilon, q_h), (x_0, \gamma_{l_0}), \ldots, (x_k, \gamma_{l_k}),$$

with $x_i \in \{1,\ldots,n\}^*$, will mean that the element $cu_{l_1}\cdots u_{l_k}$ of $\Delta^{\mathcal{U}_1}$ is homomorphically mapped to the element $a_{i_1}w_{i_2}\cdots w_{i_r}$ of $\Delta^{\mathcal{U}_2}$ for $x_k = i_1 i_2 \cdots i_r$. (Here we abuse notation, and by a *run with a prefix* of the above mentioned form, we assume a path $y_0, y_h, y_{l_0}, \ldots, y_{l_k}$ in a run tree with $\mathbf{r}(y_0) = (\epsilon, q_0)$, $\mathbf{r}(y_h) = (\epsilon, q_h)$, and $\mathbf{r}(y_{l_i}) = (x_i, \gamma_{l_i})$.) Consider now $u_{l_{k+1}}$ such that $u_{l_k} \rightsquigarrow_{\mathcal{K}_1} u_{l_{k+1}}$, then either $u_{l_k} \xrightarrow[\Sigma]{} u_{l_{k+1}}$ or $u_{l_k} \dashrightarrow_{\Sigma} u_{l_{k+1}}$. In the first case the run continues as

$$(\epsilon, q_0), (\epsilon, q_h), (x_0, \gamma_{l_0}), \ldots, (x_k, \gamma_{l_k}), (x_{k+1}, \gamma_{l_{k+1}})$$

determining the mapping of $cu_{l_1}\cdots u_{l_k}u_{l_{k+1}}$ to an immediate successor or predecessor of $a_{i_1}w_{i_2}\cdots w_{i_r}$. In the second case we enter into an "expecting state" $\chi_{l_{k+1}}$ and by our definition of the transition function, the automata will start to traverse the tree non-deterministically until it reaches some $\sigma \in \Delta^{\mathcal{U}_2}$:

$$(\epsilon, q_0), (\epsilon, q_h), (x_0, \gamma_{l_0}), \ldots, (x_k, \gamma_{l_k}), (x_{k+1}, \chi_{l_{k+1}}), \ldots, (x_{k'}, \chi_{l_{k+1}}),$$

where $x_{k'} = i_1 i_2 \cdots i_r$ for some $i_1, i_2, \ldots, i_r$, and $\sigma = a_{i_1}w_{i_2}\cdots w_{i_r}$. Then, once the desired σ is reached, the previous run should continue to $(x_{k'}, \gamma_{l_{k+1}})$ meaning that the homomorphism for $cu_{l_1}\cdots u_{l_k}u_{l_{k+1}}$ is σ. Notice that it is not yet guaranteed that a state $\chi_{l_{k+1}}$ will eventually switch to $\gamma_{l_{k+1}}$ in the run. However, by our definition of the transitions it is only possible when the run has infinitely many $\chi_{l_{k+1}}$, which would contradict to our acceptance condition presented below.

It remans to explain the purpose of the states κ_l^i. Suppose we have a run with a prefix $(\epsilon, q_0), (\epsilon, q_h), (x_0, \gamma_{l_0}), \ldots, (x_k, \gamma_{l_k})$, where $x_k = i_1$, i.e., the homomorphic image of $cu_{l_1}\cdots u_{l_k}$ is a_{i_1}, then suppose $u_{l_k} \xrightarrow[\Sigma]{} u_{l_{k+1}}$ and the mapping of $cu_{l_1}\cdots u_{l_k}u_{l_{k+1}}$ should be the constant a_j. The run therefore should proceed into the j-th successor of the parent (which is ϵ in this case) of x_k and the state $\gamma_{l_{k+1}}$, we implement this by means of an intermediate state $\kappa_{l_{k+1}}^j$ so that the following is a possible extension of the previous run

$$(\epsilon, q_0), (\epsilon, q_h), (x_0, \gamma_{l_0}), \ldots, (x_k, \gamma_{l_k}), (\epsilon, \kappa_{l_{k+1}}^j), (j, \gamma_{l_{k+1}}).$$

The transitions for the states of Q_h are now defined as follows, where $1 \leq k \leq m$, $1 \leq j \leq n_w$, and $t \in \{a_1, \ldots, a_{n_a}, w_1, \ldots, w_{n_w}\}$:

$$\delta(q_h, L) = \begin{cases} (i, \gamma_0), & \text{if } L = R \text{ and } c = a_i \text{ for some } i, \\ \bot, & \text{otherwise;} \end{cases} \tag{6}$$

$$\delta(\chi_k, \bar{t}) = (0, \gamma_k) \vee \bigvee_{t \rightsquigarrow_{\mathcal{K}_2} w_j} (j, \chi_k) \vee (-1, \chi_k); \tag{7}$$

$$\delta(\chi_k, R) = \bigvee_{i=1}^{n_a} (i, \chi_k); \tag{8}$$

$$\text{for } 1 \le i \le n_a, \quad \delta(\kappa_k^i, L) = \begin{cases} (i, \gamma_k), & \text{if } L = R, \\ \bot, & \text{otherwise;} \end{cases} \tag{9}$$

$$\text{for } q \in Q_h, \quad \delta(q, S) = \bot. \tag{10}$$

Finally, $\delta(\gamma_k, \overline{a}) =$

$$\tau_a^{u_k} \wedge \bigwedge_{u_k \dashrightarrow_{\Sigma} u_l} (0, \chi_l) \wedge \bigwedge_{u_k \xrightarrow{\Sigma} u_l} \Big(\bigvee_{a \rightsquigarrow_{\mathcal{K}_2} w_j} [\rho_{a,w_j}^{u_k,u_l} \wedge (j, \gamma_l)] \vee \bigvee_{i=1}^{n_a} [\rho_{a,a_i}^{u_k,u_l} \wedge (-1, \kappa_l^i)] \Big), \tag{11}$$

and $\delta(\gamma_k, \overline{w}) =$

$$\tau_w^{u_k} \wedge \bigwedge_{u_k \dashrightarrow_{\Sigma} u_l} (0, \chi_l) \wedge \bigwedge_{u_k \xrightarrow{\Sigma} u_l} \Big(\bigvee_{w \rightsquigarrow_{\mathcal{K}_2} w_j} [\rho_{w,w_j}^{u_k,u_l} \wedge (j, \gamma_l)] \vee [\eta_w^{u_k,u_l} \wedge (-1, \gamma_l)] \Big), \tag{12}$$

for $1 \le l \le m$, $a \in \{a_1, \ldots, a_{n_a}\}$, $w \in \{w_1, \ldots, w_{n_w}\}$, and $0 \le k \le m$ in (11), $1 \le k \le m$ in (12).

For the *acceptance condition* we take $F = \{\gamma_i \mid 1 \le i \le m\}$. Observe that neither the states ω_i of Q_f nor χ_k of Q_h are in F. This implies that a tree having an infinite branch of $\overline{w}_i$ labels, or a tree having a run in which the mapping of the disconnected successor as $u_{l_{k+1}}$ (such that $u_{l_k} \dashrightarrow_{\Sigma} u_{l_{k+1}}$) in the example above is "infinitely postponed" will be rejected. On the other hand, each accepted tree represents some finite substructure of $\mathcal{U}_2$ to which $\mathcal{U}_1^c$ can be Σ-homomorphically mapped.

We prove that verifying whether $\mathcal{U}_1$ can be Σ-homomorphically mapped to $\mathcal{U}_2^D$ for some finite $D \subseteq \Delta^{\mathcal{U}_2}$ reduces to checking the non-emptiness problem of $\mathbb{A}_c$.

Lemma 5.3.6. *Let $\mathcal{K}_1, \mathcal{K}_2$ be KBs and Σ a signature. There exists a finite subset D of $\Delta^{\mathcal{U}_2}$ and a Σ-homomorphism from $\mathcal{U}_1$ to $\mathcal{U}_2^D$ if and only if*

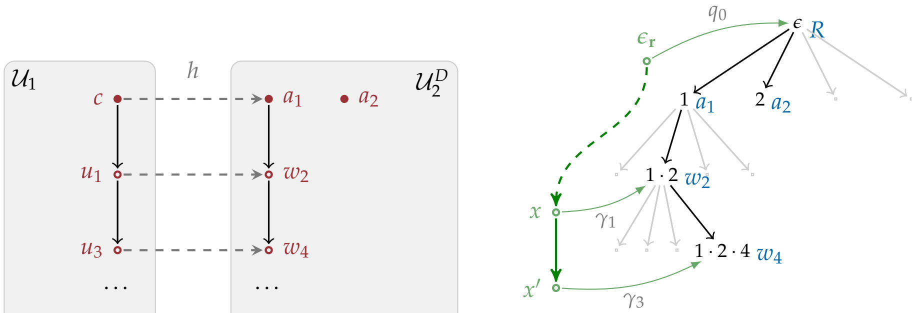
(ABOX) $\mathbf{r}_\Sigma^{\mathcal{U}_1}(a, b) \subseteq \mathbf{r}_\Sigma^{\mathcal{U}_2}(a, b)$, *for all $a, b \in \mathsf{Ind}(\mathcal{K}_1)$;*

(AUT) *the language of the automata $\mathbb{A}_c$ is non-empty, for each $c \in \mathsf{Ind}(\mathcal{K}_1)$.*

Proof. ($\Rightarrow$) Let $D \subseteq \Delta^{\mathcal{U}_2}$ be a finite set, and h a Σ-homomorphism from $\mathcal{U}_1$ to $\mathcal{U}_2^D$. We construct a labeled tree $T = (\{1, \ldots, n\}^*, V)$ where $n = \max(n_a, n_w)$ and show that $T \in \mathcal{L}(\mathbb{A}_c)$, for each $c \in \mathsf{Ind}(\mathcal{K}_1)$. The labeling function V is defined as follows:

$$\begin{aligned} &V(\epsilon) = R, \\ &V(i) = \overline{a}_i, && \text{for each } a_i \in D \cap \mathsf{Ind}(\mathcal{K}_2) \\ &V(i_1 i_2 \cdots i_r) = \overline{w}_{i_r}, && \text{for each } a_{i_1} w_{i_2} \cdots w_{i_r} \in D \\ &V(x) = S, && \text{for each } x \in \{1, \ldots, n\}^* \text{ s.t. } V(x) \text{ was not defined above.} \end{aligned}$$

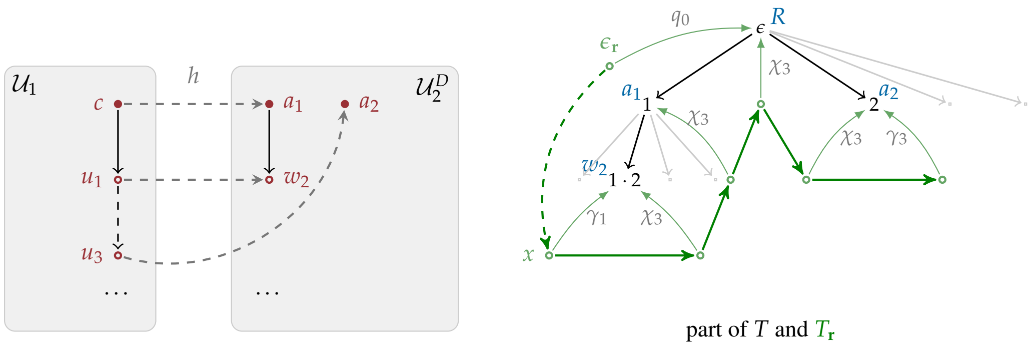
To show that $T \in \mathcal{L}(\mathbb{A}_c)$, we construct a run tree $(T_\mathbf{r}, \mathbf{r})$ of $\mathbb{A}_c$ on T. The idea behind this constructions is the following. Assume $x \in T_\mathbf{r}$ with $\mathbf{r}(x) = (y, q)$, $y \in \{1, \ldots, n\}^*$, and $V(y) = L$. To satisfy the transition function $\delta(q, L)$, which can be viewed as a conjunction of formulas Φ_i, $\delta(q, L) = \bigwedge_i \Phi_i$, where each Φ_i is a disjunction of "simple" formulas ψ_j^i, $\Phi_i = \bigvee_j \psi_j^i$, we construct exactly one child for each Φ_i, that is, we pick exactly one ψ_j^i from Φ_i. To choose the exact ψ_j^i we are making use of the given homomorphism h. Thus, for instance, if $\mathbf{r}(x) = (1 \cdot 2, \gamma_1)$, $V(1 \cdot 2) = \overline{w}_2$, the current path in $\mathcal{U}_1$ is cu_1 (this path can be obtained from the path from the root of $T_{\mathbf{r}}$ to x), $h(cu_1) = a_1w_2$, and $u_1 \xrightarrow[\Sigma]{} u_3$ and $h(cu_1u_3) = a_1w_2w_4$, then we satisfy $\psi_j^i = (4, \gamma_3)$, so x would have a child x' with $\mathbf{r}(x') = (1 \cdot 2 \cdot 4, \gamma_3)$.



part of T and $T_{\mathbf{r}}$

And if, instead, $u_1 \dashrightarrow_{\Sigma} u_3$ and $h(cu_1u_3) = a_2$, we switch to the "expecting" state χ_3 and remain in this state while traversing the tree $\{1, \ldots, n\}^*$ from the node $1 \cdot 2$ via the root to the node 2. Once node 2 is reached, we switch to the state γ_3. The choices for satisfying the transition function follow from that. Thus, a run from x will be as follows:

$$(1 \cdot 2, \gamma_1), (1 \cdot 2, \chi_3), (1, \chi_3), (\epsilon, \chi_3), (2, \chi_3), (2, \gamma_3).$$



part of T and $T_{\mathbf{r}}$

More formally, the tree structure $T_{\mathbf{r}}$ and the labeling function $\mathbf{r}$ are defined inductively as follows, where for $(x, q) \in \{1, \ldots, n\}^* \times Q$, $f((x, q))$ denotes x, and $(z)^q$ denotes $z \cdots z$, where z is repeated q times:

- $\epsilon \in T_{\mathbf{r}}$ is the root of $T_{\mathbf{r}}$ and $\mathbf{r}(\epsilon) = (\epsilon, q_0)$,
- ϵ has two children 0_f and 0_h such that $\mathbf{r}(0_f) = (\epsilon, q_f)$ and $\mathbf{r}(0_h) = (\epsilon, q_h)$,
- 0_f has children $c_1, \ldots, c_{n_a}$ such that $\mathbf{r}(c_i) = (i, \alpha_i)$,
- for $i \in \{1, \ldots, n_a\}$ and each w_j such that $a_i \rightsquigarrow_{\mathcal{K}_2} w_j$, c_i has a child $c_i \cdot w_j$ with $\mathbf{r}(c_i \cdot w_j) = (i \cdot j, \omega_j)$,
- for each node in $T_{\mathbf{r}}$ of the form $x = c_{i_1}w_{i_2} \cdots w_{i_r}$, such that $r \geq 2$ and $a_{i_1}w_{i_2} \cdots w_{i_r} \in D$, and each w_j such that $w_{i_r} \rightsquigarrow_{\mathcal{K}_2} w_j$, x has a child $x \cdot w_j$ with $\mathbf{r}(x \cdot w_j) = (i_1 i_2 \cdots i_r j, \omega_j)$,
- 0_h has one child y_0 with $\mathbf{r}(y_0) = (i, \gamma_0)$ where $i \in \{1, \ldots, n_a\}$ is such that $c = a_i$,
- for each node of the form $x = y_0 \cdot (z_{l_1})^{q_1} \cdot y_{l_1} \cdot (z_{l_2})^{q_2} \cdot y_{l_2} \cdots (z_{l_k})^{q_k} \cdot y_{l_k}$, where $k \geq 0$, $q_i \geq 0$, z_{l_i} denotes x_{l_i} or $k_{l_i}^j$, and $f(\mathbf{r}(x)) = j_1 \cdots j_s$ with $s \geq 1$, and for each u_l such that $u_{l_k} \dashrightarrow_{\Sigma} u_l$ and $h(cu_{l_1} \cdots u_{l_k}u_l) = a_{i_1}w_{i_2} \cdots w_{i_r}$,
  - x has a child $x \cdot x_l$ with $\mathbf{r}(x \cdot x_l) = (j_1 \cdots j_s, \chi_l)$;

- if $j_1 = i_1$, let t be the number s.t. $j_1 = i_1, \ldots, j_t = i_t$, and
  if $j_1 \neq i_1$, let $t = 0$, then
  * every node of the form $x' = x(x_l)^q$, $1 \leq q \leq s-t$, has one child $x' \cdot x_l$ with $\mathbf{r}(x' \cdot x_l) = (j_1 \cdots j_{s-q}, \chi_l)$,
  * every node of the form $x' = x(x_l)^q$, $s-t+1 \leq q \leq s-t+r-t$ has one child $x' \cdot x_l$ with $\mathbf{r}(x' \cdot x_l) = (j_1 \cdots j_t i_{t+1} \cdots i_{t+q-(s-t)}, \gamma_l)$, and
  * node $x' = x(x_l)^{s-t+r-t+1}$ has one child $x' \cdot y_l$ with $\mathbf{r}(x' \cdot y_l) = (i_1 \cdots i_r, \gamma_l)$.

- for each node of the form $x = y_0 \cdot (z_{l_1})^{q_1} \cdot y_{l_1} \cdot (z_{l_2})^{q_2} \cdot y_{l_2} \cdots (z_{l_k})^{q_k} \cdot y_{l_k}$, where $k \geq 0$, $q_i \geq 0$, z_{l_i} denotes x_{l_i} or $k^j_{l_i}$, and $f(\mathbf{r}(x)) = i$, and for each u_l such that $u_{l_k} \xrightarrow[\Sigma]{} u_l$, x has a child
  - $x \cdot y_l$ with $\mathbf{r}(x \cdot y_l) = (i \cdot j, \gamma_l)$, if $h(cu_{l_1} \cdots u_{l_k} u_l) = a_i w_j$,
  - $x \cdot k^j_l$ with $\mathbf{r}(x \cdot k^j_l) = (\epsilon, \kappa^j_l)$, if $h(cu_{l_1} \cdots u_{l_k} u_l) = a_j$.
- for each node of the form $x = y_0 \cdot (z_{l_1})^{q_1} \cdot y_{l_1} \cdot (z_{l_2})^{q_2} \cdot y_{l_2} \cdots (z_{l_k})^{q_k} \cdot y_{l_k}$, where $k \geq 0$, $q_i \geq 0$, z_{l_i} denotes x_{l_i} or $k^j_{l_i}$, and $f(\mathbf{r}(x)) = i_1 \cdots i_{r'}$ for $r' \geq 2$, and for each u_l such that $u_{l_k} \xrightarrow[\Sigma]{} u_l$ and $h(cu_{l_1} \cdots u_{l_k} u_l) = a_{i_1} w_{i_2} \cdots w_{i_r}$, x has a child $x \cdot y_l$ with $\mathbf{r}(x \cdot y_l) = (i_1 \cdots i_r, \gamma_l)$.
- for each node of the form $x = y_0 \cdot z_1 \cdots z_q \cdot k^j_l$, $q \geq 0$ and $z_i \in \{y_i, x_i, k^{i'}_i\}$, x has one child $x \cdot y_l$ with $\mathbf{r}(x \cdot y_l) = (j, \gamma_k)$.

It is easy to see that $(T_{\mathbf{r}}, \mathbf{r})$ is an accepting run of $\mathbb{A}_c$.

($\Leftarrow$) Assume that the language of $\mathbb{A}_c$ is non-empty and $T = (\{1, \ldots, n\}^*, V) \in \mathcal{L}(\mathbb{A}_c)$. Let $(T_{\mathbf{r}}, \mathbf{r})$ be an accepting run of $\mathbb{A}_c$ over T. We construct a finite set $D_c \subseteq \Delta^{\mathcal{U}_2}$ and a Σ-homomorphism h from $\mathcal{U}^c_1$ to $\mathcal{U}^{D_c}_2$ using T and $(T_{\mathbf{r}}, \mathbf{r})$.

Firstly, we prove that T encodes a finite subset of $\Delta^{\mathcal{U}_2}$. We show

(a) for each $i \in \{1, \ldots, n_a\}$, $V(i) = \overline{a}_i$;

(b) for each $k \geq 2$, such that $a_{i_1} w_{i_2} \cdots w_{i_k} \in \Delta^{\mathcal{U}_2}$, and for each $2 \leq j < k$, $V(i_1 \cdots i_j) = \overline{w}_{i_j}$, then $V(i_1 \cdots i_k) = \overline{w}_{i_k}$ or $V(i_1 \cdots i_k) = S$;

(c) for each infinite path $a_{i_1} \cdots w_{i_j} \cdots \in \Delta^{\mathcal{U}_2}$, there exists $j \geq 2$, s.t. $V(i_1 \cdots i_j) = S$.

Proof of **(a)**: by definition of $\delta(\alpha_i, L)$.

Proof of **(b)**: for the sake of contradiction, assume for some $a_{i_1} w_{i_2} \cdots w_{i_k} \in \Delta^{\mathcal{U}_2}$, $k \geq 2$, for each $2 \leq j < k$, $V(i_1 \cdots i_j) = \overline{w}_{i_j}$, but $V(i_1 \cdots i_k) = R$ or $V(i_1 \cdots i_k) = \overline{a}_i$. Since $(T_{\mathbf{r}}, \mathbf{r})$ is a run over T there exists a path in $T_{\mathbf{r}}$ of the form

$$(\epsilon, q_0), (\epsilon, q_f), (i_1, \alpha_{i_1}), (i_1 i_2, \omega_{i_2}), \ldots, (i_1 \cdots i_k, \omega_{i_k}).$$

Then by definition of the transition function, both $\delta(\omega_{i_k}, R) = \perp$ and $\delta(\omega_{i_k}, \overline{a}_i) = \perp$, which contradicts the assumption $(T_{\mathbf{r}}, \mathbf{r})$ is a run.

Proof of **(c)**: By contradiction, assume that there exists an infinite path $a_{i_1} \cdots w_{i_j} \cdots$ in $\Delta^{\mathcal{U}_2}$, such that for each $j \geq 2$, $V(i_1 \cdots i_j) \neq S$. Now, since $(T_{\mathbf{r}}, \mathbf{r})$ is a run of $\mathbb{A}_c$ over T, there must exist an infinite path π in $T_{\mathbf{r}}$ of the form

$$(\epsilon, q_0), (\epsilon, q_f), (i_1, \alpha_{i_1}), (i_1 i_2, \omega_{i_2}), \ldots, (i_1 \cdots i_j, \omega_{i_j}), \ldots.$$

Since $\mathit{inf}(\pi) \cap \{\gamma_1, \ldots, \gamma_{n_w}\} = \emptyset$ we obtain a contradiction with the assumption that $(T_{\mathbf{r}}, \mathbf{r})$ is an accepting run. Therefore, let $d \geq 2$ be the depth of S, i.e., for each

$a_{i_1} \cdots w_{i_j} \cdots \in \Delta^{\mathcal{U}_2}$, for some $j \leq d$, $V(i_1 \cdots i_j) = S$. The finite set D_c is given by $\{a_{i_1} w_{i_2} \cdots w_{i_{d-1}} \in \Delta^{\mathcal{U}_2}\}$.

Next, we show there exists a Σ-homomorphism from $\mathcal{U}_1^c$ to $\mathcal{U}_2^D$ by constructing that h. By induction of k, we build $h(cu_{l_1} \cdots u_{l_k})$ for each $cu_{l_1} \cdots u_{l_k} \in \Delta^{\mathcal{U}_1^c}$.

Base of induction. First, in $T_{\mathbf{r}}$ there must exist a path $(\epsilon, q_0), (\epsilon, q_h)$, and as $T_{\mathbf{r}}$ is a run, for some i, $c = a_i$, hence this path continues with (i, γ_0) (and the current path is $(\epsilon, q_0), (\epsilon, q_h), (i, \gamma_0))$. Then, $\delta(\gamma_0, \overline{a}_i)$ is satisfied, which means that $\tau^c_{a_i} = \top$ and, in turn, $\mathbf{t}^{\mathcal{U}_1}_\Sigma(c) \subseteq \mathbf{t}^{\mathcal{U}_2}_\Sigma(a_i)$, so we can set $h(c) = a_i$.

Inductive step. Assume h is defined for each path of length $k+1$ in $\Delta^{\mathcal{U}_1}$, $k \geq 0$, let $cu_{l_1} \cdots u_{l_k} \in \Delta^{\mathcal{U}_1^c}$ (u_{l_0} denotes c), and $h(cu_{l_1} \cdots u_{l_k}) = a_{i_0} w_{i_1} \cdots w_{i_r}$, and assume the current path π in $T_{\mathbf{r}}$ is of the form

$$(\epsilon, q_0), (\epsilon, q_h), (i_0, \gamma_0), (x, q)^*, \ldots, (i_0 \cdots i_r, \gamma_k),$$

where $(x, q)^*$ denotes a finite (possibly empty) sequence of tuples (x, q) with $x \in \{1, \ldots, n\}^*$ and $q \in \{\gamma_l, \chi_l, \kappa^i_l \mid 1 \leq l \leq m, 1 \leq i \leq n_a\}$. Then $\delta(\gamma_k, \overline{w}_{i_r})$ (recall, that $i_0 \cdots i_r \in T$ is labeled with $\overline{w}_{i_r}$) is satisfied. Now, let $u_{l_k} \rightsquigarrow_{\mathcal{K}_1} u_{l_{k+1}}$. If $u_{l_k} \xrightarrow[\Sigma]{} u_{l_{k+1}}$, then at least one of the formulas

$$\begin{array}{rcll} \psi_j & = & \rho^{u_{l_k}, u_{l_{k+1}}}_{w_{i_r}, w_j} \wedge (j, \gamma_{l_{k+1}}), & \text{for } w_{i_r} \rightsquigarrow_{\mathcal{K}_2} w_j \ (w_{i_0} \text{ denotes } a_{i_0}), \\ \psi_i & = & \rho^{u_{l_k}, u_{l_{k+1}}}_{a_{i_0}, a_i} \wedge (-1, \kappa^i_{l_{k+1}}), & \text{if } r = 0, \\ \psi_{-1} & = & \eta^{u_{l_k}, u_{l_{k+1}}}_{w_{i_r}} \wedge (-1, \gamma_{l_{k+1}}), & \text{if } r > 0, \end{array}$$

is satisfied. Assume ψ_j is satisfied for some $j \in \{1, \ldots, n_w\}$: then $\rho^{u_{l_k}, u_{l_{k+1}}}_{w_{i_r}, w_j} = \top$, hence $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u_{l_k}, u_{l_{k+1}}) \subseteq \mathbf{r}^{\mathcal{G}_2}_\Sigma(w_{i_r}, w_j)$, and the run is continued with $(i_0 \cdots i_r j, \gamma_{l_{k+1}})$. Moreover, $\delta(\gamma_{l_{k+1}}, \overline{w}_j)$ is satisfied, so $\tau^{u_{l_{k+1}}}_{w_j} = \top$, i.e., $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u_{l_{k+1}}) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(w_j)$. Therefore, we can set $h(cu_{l_1} \cdots u_{l_{k+1}})$ to be equal to $a_{i_0} w_{i_1} \cdots w_{i_r} w_j$.

In the case $r = 0$ and ψ_i is satisfied for some $i \in \{1, \ldots, n_a\}$, we have that $\rho^{u_{l_k}, u_{l_{k+1}}}_{a_{i_0}, a_i} = \top$, hence $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u_{l_k}, u_{l_{k+1}}) \subseteq \mathbf{r}^{\mathcal{G}_2}_\Sigma(a_{i_0}, a_i)$, and the run is continued with $(\epsilon, \kappa^i_{l_{k+1}}), (i, \gamma_{l_{k+1}})$. Moreover, $\delta(\gamma_{l_{k+1}}, \overline{a}_i)$ is satisfied, so $\tau^{u_{l_{k+1}}}_{a_i} = \top$, i.e., $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u_{l_{k+1}}) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(a_i)$. Therefore, we can set $h(cu_{l_1} \cdots u_{l_{k+1}})$ to be equal to a_i.

Alternatively, if for $r > 0$, ψ_{-1} is satisfied, it follows that $\eta^{u_{l_k}, u_{l_{k+1}}}_{w_{i_r}} = \top$, hence $\{R^- \mid R \in \mathbf{r}^{\mathcal{G}_1}_\Sigma(u_{l_k}, u_{l_{k+1}})\} \subseteq \mathbf{r}^{\mathcal{G}_2}_\Sigma(w_{i_{r-1}}, w_{i_r})$, and the run is continued with $(i_0 \cdots i_{r-1}, \gamma_{l_{k+1}})$. Moreover, $\delta(\gamma_{l_{k+1}}, \overline{w}_{i_{r-1}})$ is satisfied, so $\tau^{u_{l_{k+1}}}_{w_{i_{r-1}}} = \top$, i.e., $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u_{l_{k+1}}) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(w_{i_{r-1}})$. Therefore, we can set $h(cu_{l_1} \cdots u_{l_{k+1}})$ to be equal to $a_{i_0} w_{i_1} \cdots w_{i_{r-1}}$. It concludes the inductive step for the case $u_{l_k} \xrightarrow[\Sigma]{} u_{l_{k+1}}$.

Consider now, $u_{l_k} \dashrightarrow_{\Sigma} u_{l_{k+1}}$. Then the run continues with $(i_1 \cdots i_r, \chi_{l_{k+1}})$. Let

$$(x_1, \chi_{l_{k+1}}), \ldots, (x_j, \chi_{l_{k+1}}), (x_j, \gamma_{l_{k+1}})$$

be a continuation of the current path $\pi \cdot (i_1 \cdots i_r, \chi_{l_{k+1}})$ in $T_{\mathbf{r}}$, and $x_j = j_0 \cdots j_s$. Then $\delta(\gamma_{l_{k+1}}, \overline{w}_{j_s})$ is satisfied, so $\tau^{u_{l_{k+1}}}_{w_{j_s}} = \top$, and $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u_{l_{k+1}}) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(w_{j_s})$. Since $\mathbf{r}^{\mathcal{U}_1}_\Sigma(cu_{l_1} \cdots u_{l_k}, cu_{l_1} \cdots u_{l_{k+1}}) = \emptyset$, we can set $h(cu_{l_1} \cdots u_{l_{k+1}})$ to be equal to $a_{j_0} w_{j_1} \cdots w_{j_s}$.

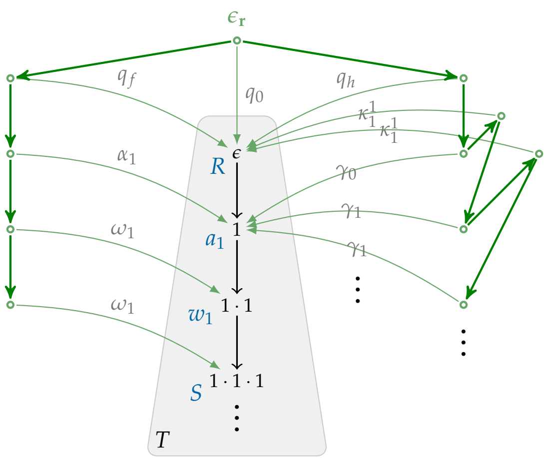
Note that the runs considered in the induction never visit a node labeled with S, otherwise it contradicts the definition of a run. Therefore, in such a manner, we can define h, a Σ-homomorphism from $\mathcal{U}_1^c$ to $\mathcal{U}_2^{D_c}$. A Σ-homomorphism from $\mathcal{U}_1$ to $\mathcal{U}_2^D$ for $D = \bigcup_c D_c$ is defined as the union of h for each $c \in \mathsf{Ind}(\mathcal{K}_1)$. $\square$

The following examples explain how the algorithm for checking the existence of a universal solution with extended ABoxes, which is based on the automata construction described above, works.

Example 5.3.7. Consider $\mathcal{M}$ and $\mathcal{K}_1$ from Example 4.1.9, i.e., $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot)\}$, $\Sigma_2 = \{Q(\cdot,\cdot)\}$, and $\mathcal{T}_{12} = \{R \sqsubseteq Q, S \sqsubseteq Q\}$, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$, where $\mathcal{A}_1 = \{A(a), S(a,a)\}$ and $\mathcal{T}_1 = \{A \sqsubseteq \exists R,\ \exists R^- \sqsubseteq \exists R\}$.

So we construct automaton $\mathbb{A}_a$ for $\mathcal{K}$, $\mathcal{K}'$ and Σ, where $\mathcal{K} = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, $\mathcal{K}' = \mathcal{K}$ and $\Sigma = \Sigma_2$. Moreover, $\mathsf{Ind}(\mathcal{K}') = \{a_1\}$, $\mathsf{Wit}(\mathcal{K}') = \{w_1\}$ and $\mathsf{Wit}(\mathcal{K}) = \{u_1\}$, where $a_1 = a$, $w_1 = w_R$, and $u_1 = w_R$. Thus $n = 1$ and $\mathbb{A}_a$ accepts trees of the form $(\{1\}^*, V)$ and $V \in \{R, S, a_1, w_1\}$ with the set of accepting states $F = \{\gamma_1\}$.

Below we depict a tree $T \in \mathcal{L}(\mathbb{A}_a)$ with an accepting run over T which starts in $\epsilon_\mathbf{r}$ with $\mathbf{r}(\epsilon_\mathbf{r}) = (\epsilon, q_0)$.

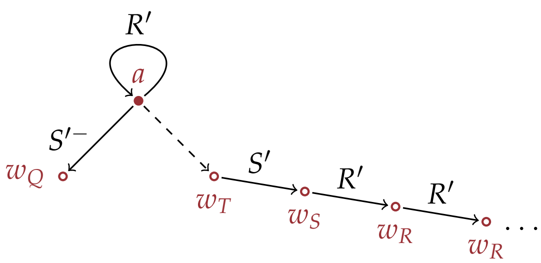


From T we can extract an ABox $\mathcal{A}_2 = \{Q(a,a), Q(a, n_{1\cdot 1})\}$, which is also a (non-minimal) universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. ■

Example 5.3.8. Consider mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, and source TBox $\mathcal{T}_1$ and ABox $\mathcal{A}_1$, where

$$\begin{aligned}
\Sigma_1 &= \{P(\cdot,\cdot),\ Q(\cdot,\cdot),\ T(\cdot,\cdot),\ S(\cdot,\cdot),\ R(\cdot,\cdot)\}, \quad \Sigma_2 = \{S'(\cdot,\cdot),\ R'(\cdot,\cdot)\},\\
\mathcal{T}_{12} &= \{Q \sqsubseteq S'^-,\ P \sqsubseteq R',\ S \sqsubseteq S',\ R \sqsubseteq R'\},\\
\mathcal{A}_1 &= \{\exists T(a),\ \exists Q(a),\ P(a,a)\},\\
\mathcal{T}_1 &= \{\exists T^- \sqsubseteq \exists S,\ \exists S^- \sqsubseteq \exists R,\ \exists R^- \sqsubseteq \exists R\}.
\end{aligned}$$

The projection of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ over Σ_2 is shown below:

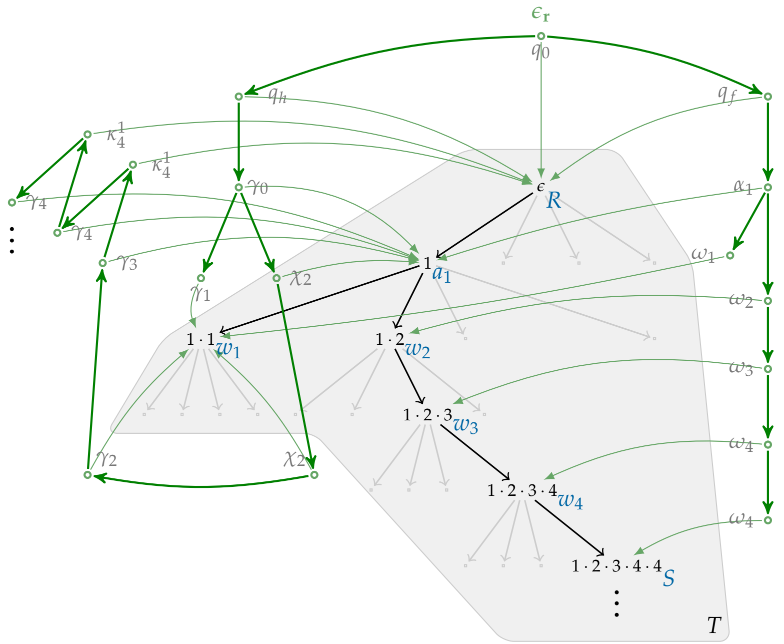


We construct automaton $\mathbb{A}_a$ for $\mathcal{K}_1$, $\mathcal{K}_2$ and Σ, where $\mathcal{K}_1 = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, $\mathcal{K}_2 = \mathcal{K}_1$, and $\Sigma = \Sigma_2$. Then $\mathsf{Ind}(\mathcal{K}_2) = \{a_1\}$, $\mathsf{Wit}(\mathcal{K}_2) = \{w_1, w_2, w_3, w_4\}$ and $\mathsf{Wit}(\mathcal{K}_1) = \{u_1, u_2, u_3, u_4\}$, where $a_1 = a$, and

$$w_1 = u_1 = w_Q, \quad w_2 = u_2 = w_T, \quad w_3 = u_3 = w_S, \quad w_4 = u_4 = w_R.$$

Thus $n = 4$, so $\mathbb{A}_a$ accepts trees of the form $(\{1, \dots, 4\}^*, V)$, $V \in \{R, S, a_1, w_1, w_2, w_3, w_4\}$, and the set of accepting states $F = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$.

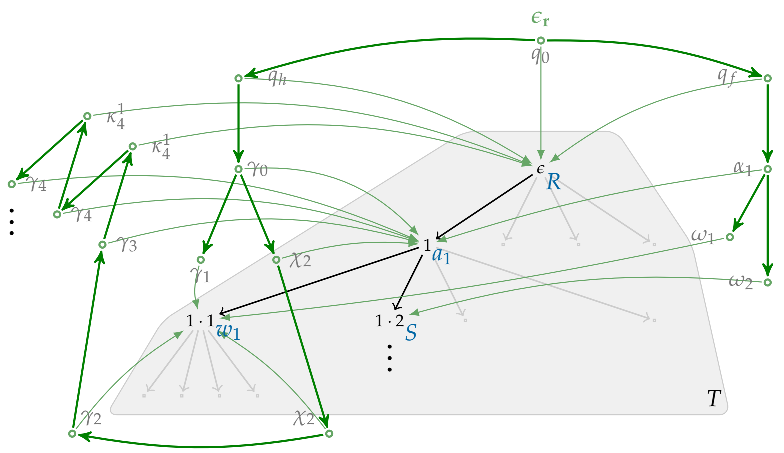
Then we check for non-emptiness of $\mathbb{A}_a$. It turns out to be non-empty, so below we depict a tree $T \in \mathcal{L}(\mathbb{A}_a)$ with an accepting run over T which starts in $\epsilon_\mathbf{r}$ with $\mathbf{r}(\epsilon_\mathbf{r}) = (\epsilon, q_0)$. Here, the nodes of T that are never visited by $\mathbb{A}_a$ are depicted in gray as they are not relevant.



From T we can extract a universal solution

$$\mathcal{A}_2 = \{R'(a, a),\ S'^{-}(a, n_{1\cdot1}),\ S'(n_{1\cdot2}, n_{1\cdot2\cdot3}),\ R'(n_{1\cdot2\cdot3}, n_{1\cdot2\cdot3\cdot4})\}$$

for $\mathcal{K}_1$ under $\mathcal{M}$. Observe that $\mathcal{A}_2$ is not a minimal universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. Instead, a tree T' and a run that correspond to the minimal universal solution $\mathcal{A}_2' = \{R'(a, a),\ S'(a, n_{1\cdot1})\}$ can be depicted as follows:



■

It is well known that the non-emptiness problem for 2ATA is solvable in EXPTIME. Then, since the automaton $\mathbb{A}_c$ we constructed above is of polynomial size, summing up, we get:

Theorem 5.3.9. *The non-emptiness problem for extended universal solutions is* PSPACE-*hard and in* EXPTIME.*For a given mapping $\mathcal{M}$ and a given KB $\mathcal{K}_1$, if a universal solution $\mathcal{A}_2$ exists, then it is at most exponentially large in the size of $\mathcal{K}_1 \cup \mathcal{M}$.*

5.4 UNIVERSAL SOLUTIONS IN DL-LITE$_{RDFS}$

The differences of *DL-Lite*$_{RDFS}$ with *DL-Lite*$_{\mathcal{R}}$ is that *DL-Lite*$_{RDFS}$ axioms cannot have either existential quantification, or negated concepts or roles on the right-hand side. It implies that for a *DL-Lite*$_{RDFS}$ KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, the canonical model $\mathcal{U}_{\mathcal{K}}$ of $\mathcal{K}$ is, first, finite, and, second, $\Delta^{\mathcal{U}_{\mathcal{K}}}$ consists only of the constants in $\mathcal{A}$, hence $|\mathcal{U}_{\mathcal{K}}|$ is polynomial in the size of $\mathcal{K}$. Then, for each *DL-Lite*$_{RDFS}$ mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and each *DL-Lite*$_{RDFS}$ KB $\mathcal{K}_1$ over Σ_1, $\mathcal{K}_1$ is trivially Σ_2-safe with respect to $\mathcal{M}$, therefore, by Lemma 5.1.6 it follows that there exists a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$, and it is polynomially large in the size of $\mathcal{K}_1 \cup \mathcal{M}$. Thus, we obtain a trivial complexity bound for the non-emptiness problem for universal solutions in *DL-Lite*$_{RDFS}$, independently of whether simple or extended ABoxes are allowed in the target.

Theorem 5.4.1. *In DL-Lite*$_{RDFS}$*, the non-emptiness problem for universal solutions is in* TRIVIAL.

Let us turn now to the membership problem. By Lemma 5.1.2 and by the way homomorphisms are defined on constants, it follows that a simple target ABox $\mathcal{A}_2$ is a universal solution for a source KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ if and only if $\mathcal{U}_{\mathcal{A}_2}$ agrees with $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ on concepts and roles from Σ_2. The latter can be checked very efficiently, so we obtain the following complexity bound.

Theorem 5.4.2. *In DL-Lite*$_{RDFS}$*, the membership problem for simple universal solutions is* NLOGSPACE-*complete.*

Proof. We show the lower bound by reduction from the reachability problem in directed graphs. □

As for the membership problem with extended ABoxes, clearly the upper bound is inherited from the same problem in *DL-Lite*$_{\mathcal{R}}$. To see that the lower bound applies as well, observe that in the reduction from the 3-colorability problem to show that the membership problem with extended ABoxes is NP-hard we use only role inclusions, hence the instance of KB exchange problem we construct is in *DL-Lite*$_{RDFS}$. The theorem below is a straightforward corollary of Theorem 5.3.3.

Theorem 5.4.3. *In DL-Lite*$_{RDFS}$*, the membership problem for extended universal solutions is* NP-*complete.*

6 REASONING ABOUT UNIVERSAL UCQ-SOLUTIONS

In this chapter, we study universal UCQ-solutions. First in Section 6.1 we present a characterization of universal UCQ-solutions involving finite homomorphisms. Then in Section 6.2 we investigate universal UCQ-solutions with simple ABoxes in the target, and in Section 6.3 we investigate universal UCQ-solutions with extended ABoxes in the target. For the latter we essentially provide a PSPACE lower bound for the non-emptiness problem. We conclude with a complete picture for universal UCQ-solutions in the case of *DL-Lite*$_{RDFS}$ in Section 6.4, which coincides with the complexity of universal solutions in *DL-Lite*$_{RDFS}$.

6.1 CHARACTERIZATION OF UNIVERSAL UCQ-SOLUTIONS

We start with a lemma that shows that the canonical model can be used for checking certain answers to UCQs.

Lemma 6.1.1. *Let $\mathcal{K}$ be a consistent KB, $q(\vec{x})$ a* UCQ *and $\vec{a} \subseteq N_a$ a tuple of constants. Then it holds $\mathcal{K} \models q[\vec{a}]$ iff $\mathcal{U}_\mathcal{K} \models q[\vec{a}]$.*

Proof. ($\Rightarrow$) Assume $\mathcal{K} \models q[\vec{a}]$. Then for each model $\mathcal{I}$ of $\mathcal{K}$, we have that $\mathcal{I} \models q[\vec{a}]$. Since $\mathcal{U}_\mathcal{K}$ is a model of $\mathcal{K}$, it follows $\mathcal{U}_\mathcal{K} \models q[\vec{a}]$.

($\Leftarrow$) Let $\mathcal{U}_\mathcal{K} \models q[\vec{a}]$, moreover assume $\vec{a} = (a_1, \dots, a_k)$ for $a_i \in N_a$, and $q(\vec{x}) = \exists y_1 \dots \exists y_m.\phi(x_1, \dots, x_k, y_1, \dots, y_m)$. Then it follows that there exist $\sigma_1, \dots, \sigma_m \in \Delta^{\mathcal{U}_\mathcal{K}}$ such that $\mathcal{U}_\mathcal{K} \models \phi[a_1, \dots, a_k, \sigma_1, \dots, \sigma_m]$.

Let $\mathcal{I}$ be a model of $\mathcal{K}$, we show that $\mathcal{I} \models q[\vec{a}]$. By Theorem 2.1.5, there exists a homomorphism h from $\mathcal{U}_\mathcal{K}$ to $\mathcal{I}$. Then it is easy to see that

$$\mathcal{I} \models \phi[a_1, \dots, a_k, h(\sigma_1), \dots, h(\sigma_m)].$$

As $\mathcal{I}$ was an arbitrary model of $\mathcal{K}$, it follows that $\mathcal{K} \models q[\vec{a}]$. $\square$

Then we define the notion of Σ-query entailment studied in [77] that generalizes the concept of UCQ-solutions.

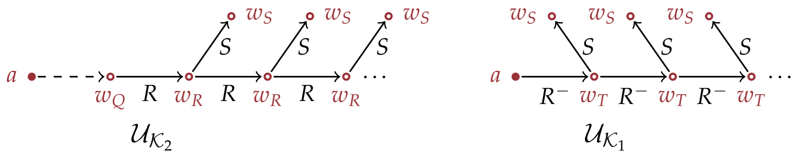
Definition 6.1.2. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be KBs, and Σ a signature. Then, $\mathcal{K}_1$* Σ-query entails *$\mathcal{K}_2$ if for each* UCQ *q over Σ,*

$$cert(q, \mathcal{K}_2) \subseteq cert(q, \mathcal{K}_1).$$

Moreover, $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-query equivalent, *if they Σ-query entail each other.*

It is well known that homomorphisms preserve answers to UCQs [72, 1], in particular, if $\mathcal{U}_{\mathcal{K}_2}$ is Σ-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_1}$, then $\mathcal{K}_1$ Σ-entails $\mathcal{K}_2$. However, it is not a necessary condition, as demonstrated by the following example.

Example 6.1.3. Consider $\Sigma = \{R(\cdot,\cdot), S(\cdot,\cdot)\}$, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \{\exists T(a)\}\rangle$ and $\mathcal{K}_2 = \langle \mathcal{T}_2, \{\exists Q(a)\}\rangle$, where $\mathcal{T}_1 = \{\exists T^- \sqsubseteq \exists T,\ \exists T^- \sqsubseteq \exists S,\ T \sqsubseteq R^-\}$ and $\mathcal{T}_2 = \{\exists Q^- \sqsubseteq \exists R,\ \exists R^- \sqsubseteq \exists R,\ \exists R^- \sqsubseteq \exists S\}$. Then one can notice that $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$ even though $\mathcal{U}_{\mathcal{K}_2}$ is not Σ-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_1}$.



■

The example above suggests that for a characterization of Σ-query entailment one has to consider finite Σ-homomorphisms. Here, given interpretations $\mathcal{I}$ and $\mathcal{J}$, we say $\mathcal{I}$ is *finitely* Σ*-homomorphically embeddable* into $\mathcal{J}$, if for every finite sub-interpretation $\mathcal{I}'$ of $\mathcal{I}$, there exists a Σ-homomorphism from $\mathcal{I}'$ to $\mathcal{J}$.

Lemma 6.1.4 ([77]). *Let $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1\rangle$ and $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2\rangle$ be consistent KBs with the corresponding canonical models $\mathcal{U}_1$ and $\mathcal{U}_2$, and Σ a signature. Then $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$ iff $\mathcal{U}_2$ is finitely Σ-homomorphically embeddable into $\mathcal{U}_1$.*

Proof. ($\Rightarrow$) Assume $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$. Let D be a finite subset of $\Delta^{\mathcal{U}_2}$ such that $D = \{a_1, \ldots, a_k, \sigma_1, \ldots, \sigma_m\}$ with $a_i \in \mathsf{Ind}(\mathcal{A}_2)$. Consider a CQ $q = \exists y_1 \ldots \exists y_m.\phi$, where for $i, i' \in \{1, \ldots, k\}$ and $j, j' \in \{1, \ldots, m\}$

$$\begin{aligned}\phi \quad = \quad & \bigwedge_{A \in \mathbf{t}_\Sigma^{\mathcal{U}_2}(a_i)} A(a_i) \ \wedge \bigwedge_{R \in \mathbf{r}_\Sigma^{\mathcal{U}_2}(a_i, a_{i'})} R(a_i, a_{i'}) \ \wedge \bigwedge_{R \in \mathbf{r}_\Sigma^{\mathcal{U}_2}(a_i, \sigma_j)} R(a_i, y_j) \ \wedge \\ & \bigwedge_{A \in \mathbf{t}_\Sigma^{\mathcal{U}_2}(\sigma_j)} A(y_j) \ \wedge \bigwedge_{R \in \mathbf{r}_\Sigma^{\mathcal{U}_2}(\sigma_j, \sigma_{j'})} R(y_j, y_{j'})\end{aligned}$$

Clearly, $\mathcal{U}_2 \models q$, as $\mathcal{U}_2 \models \phi[\sigma_1, \ldots, \sigma_m]$. Then using Lemma 6.1.1, we obtain that $\mathcal{U}_1 \models q$, therefore for some $\sigma'_1, \ldots, \sigma'_m \in \Delta^{\mathcal{U}_1}$, $\mathcal{U}_1 \models \phi[\sigma'_1, \ldots, \sigma'_m]$. We define a function $h : D \to \Delta^{\mathcal{U}_1}$ as $h(a_i) = (a_i)$ and $h(\sigma_i) = \sigma'_i$. This function is a homomorphism: it maps every constant to itself, and from $\mathcal{U}_1 \models \phi[\sigma'_1, \ldots, \sigma'_m]$ it follows that for each $d, d' \in D$, $\mathbf{t}_\Sigma^{\mathcal{U}_2}(d) \subseteq \mathbf{t}_\Sigma^{\mathcal{U}_1}(h(d))$ and $\mathbf{r}_\Sigma^{\mathcal{U}_2}(d, d') \subseteq \mathbf{r}_\Sigma^{\mathcal{U}_1}(d, d')$

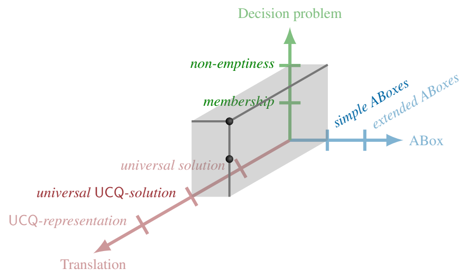
($\Leftarrow$) Assume $\mathcal{U}_2$ is finitely Σ-homomorphically embeddable into $\mathcal{U}_1$. Let q be a Σ-query such that $\mathcal{K}_2 \models q$ and $q = \exists y_1, \ldots, \exists y_m.\phi$, where ϕ is a conjunction of atoms over constants $a_1, \ldots, a_k$ and variables $y_1, \ldots, y_m$. Then $\mathcal{U}_2 \models \phi[\sigma_1, \ldots, \sigma_m]$ for some $\sigma_j \in \Delta^{\mathcal{U}_2}$. Let $D = \{a_1, \ldots, a_k, \sigma_1, \ldots, \sigma_k\}$ and h a Σ-homomorphism from $\mathcal{U}_2^D$ to $\Delta^{\mathcal{U}_1}$ with $h(a_i) = a_i$ for $i = 1, \ldots, k$. By definition of homomorphism, we have that for each concept A over Σ and $d \in D$, if $d \in A^{\mathcal{U}_2}$, then $h(d) \in A^{\mathcal{U}_1}$, and for each role R over Σ and $d, d' \in D$, if $(d, d') \in R^{\mathcal{U}_2}$, then $(h(d), h(d')) \in R^{\mathcal{U}_1}$. Which in turns implies that $\mathcal{U}_1 \models \phi[h(\sigma_1), \ldots, h(\sigma_m)]$, hence, $\mathcal{K}_1 \models q$. □

As a corollary of the lemma above, we obtain a characterization of universal UCQ-solutions.

Corollary 6.1.5. *A KB $\mathcal{K}_2$ over Σ_2 is a universal* UCQ*-solution for a KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1\rangle$ under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ iff $\mathcal{U}_{\mathcal{K}_2}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}\rangle, \mathcal{A}_1}$.*

6.2 UNIVERSAL UCQ-SOLUTIONS WITH SIMPLE ABOXES

In this section we study the membership and non-emptiness problems for universal UCQ-solutions with simple ABoxes. Section 6.2.1 is dedicated to the membership problem, for which we provide a PSPACE lower bound and develop an EXPTIME algorithm based on an involved reduction to reachability games. In Section 6.2.2, we show how to solve the non-emptiness problem in EXPTIME using the algorithm devised for the membership problem.



6.2.1 *The membership problem*

We start by showing that reasoning about universal UCQ-solutions is harder than reasoning about universal solutions, which can be explained by the fact that TBoxes have bigger impact on the structure of universal UCQ-solutions rather than of universal solutions. In fact, by using a reduction from the validity problem for quantified Boolean formulas, similar to a reduction in [77], we are able to prove the following:

Theorem 6.2.1. *The membership problem for universal* UCQ*-solutions is* PSPACE*-hard.*

Proof. The proof is by reduction of the validity problem for quantified Boolean formulas, known to be PSPACE-complete. Suppose we are given a QBF

$$\phi = \mathsf{Q}_1 X_1 \ldots \mathsf{Q}_n X_n \bigwedge_{j=1}^{m} C_j$$

where $\mathsf{Q}_i \in \{\forall, \exists\}$ and C_j, $1 \leq j \leq m$, are clauses over the variables X_i, $1 \leq i \leq n$.

Let $\Sigma_1 = \{A, Y_i^k, X_i^k, S_l, T_l, Q_i^k, P_i^k, R_j \mid 1 \leq j \leq m, 1 \leq i \leq n, 0 \leq l \leq n, k \in \{0,1\}\}$ where A, Y_i^k, X_i^k are concept names and the rest are role names. Let $\mathcal{T}_1$ be the following TBox over Σ_1 for $1 \leq j \leq m$, $1 \leq i \leq n$ and $k \in \{0,1\}$:

$$\begin{array}{llll}
A \sqsubseteq \exists S_0^- & \exists S_{i-1}^- \sqsubseteq \exists Q_i^k \text{ if } \mathsf{Q}_i = \forall & Q_i^k \sqsubseteq S_i & \exists (Q_i^k)^- \sqsubseteq Y_i^k \\
 & \exists S_{i-1}^- \sqsubseteq \exists S_i \text{ if } \mathsf{Q}_i = \exists & \exists S_n^- \sqsubseteq \exists R_j & \exists R_j^- \sqsubseteq \exists R_j
\end{array}$$

$$\begin{array}{lll}
A \sqsubseteq \exists T_0^- & \exists T_{i-1}^- \sqsubseteq \exists P_i^k & X_i^0 \sqsubseteq \exists R_j \text{ if } \neg X_i \in C_j \\
P_i^k \sqsubseteq T_i & \exists (P_i^k)^- \sqsubseteq X_i^k & X_i^1 \sqsubseteq \exists R_j \text{ if } X_i \in C_j
\end{array}$$

and $\mathcal{A}_1 = \{A(a)\}$.

Further, let $\Sigma_2 = \{A', Z_i^0, Z_i^1, S', R_j', P_i^{k'}, T_l'\}$ where A', Z_i^0, Z_i^1 are concept names and $S', R_j', P_i^{k'}, T_l', R_j^{l'}$ are role names, $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, and $\mathcal{T}_{12}$ the following set of inclusions:

$$\begin{array}{lllll}
A \sqsubseteq A' & S_i \sqsubseteq S' & Y_i^k \sqsubseteq Z_i^k & R_j \sqsubseteq R_j' & P_i^k \sqsubseteq P_i^{k'} \\
 & T_i \sqsubseteq S' & X_i^k \sqsubseteq Z_i^k & T_i \sqsubseteq R_j'^- & T_l \sqsubseteq T_l'
\end{array}$$

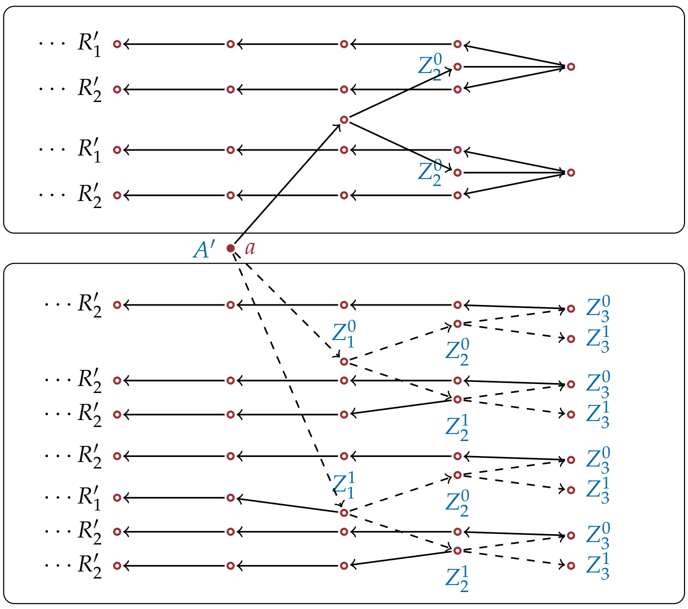
Finally, let $\mathcal{A}_2 = \{A'(a)\}$, and $\mathcal{T}_2$ the following target TBox for $1 \leq j \leq m$, $1 \leq i \leq n$ and $k \in \{0,1\}$:

$$\begin{array}{ccc}
A' \sqsubseteq \exists T_0'^{-} & \exists T_{i-1}'^{\,-} \sqsubseteq \exists P_i^{k\prime} & Z_i^0 \sqsubseteq \exists R_j' \text{ if } \neg X_i \in C_j \\
P_i^{k\prime} \sqsubseteq T_i' & \exists (P_i^{k\prime})^- \sqsubseteq Z_i^k & Z_i^1 \sqsubseteq \exists R_j' \text{ if } X_i \in C_j \\
T_i' \sqsubseteq S' & T_i' \sqsubseteq R_j'^{-} & \exists R_j'^{-} \sqsubseteq \exists R_j'
\end{array}$$

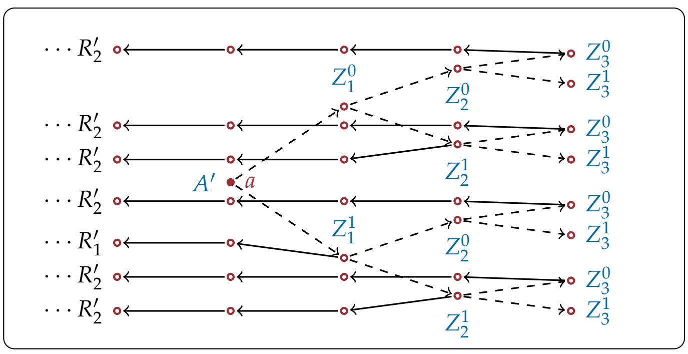
We verify that $\models \phi$ if and only if $\langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is a universal UCQ-solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$. Due to Corollary 6.1.5, it suffices to show that $\models \phi$ iff $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_2, \mathcal{A}_2 \rangle}$. The rest of the proof is similar to Lemma 5.3.4. $\square$

We illustrate the reduction by an example.

Example 6.2.2. For $\phi = \exists X_1 \forall X_2 \exists X_3 (X_1 \wedge (X_2 \vee \neg X_3))$, the projection over Σ_2 of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ and $\mathcal{U}_{\langle \mathcal{T}_2, \mathcal{A}_2 \rangle}$ can be depicted as in Figure 11a and Figure 11b, respectively. Here each edge $\longrightarrow$ is labeled with S', each edge $\dashrightarrow$ is labeled with $S', R_j'^{-}$ for



(a) The projection over Σ_2 of $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$.



(b) $\mathcal{U}_{\langle \mathcal{T}_2, \mathcal{A}_2 \rangle}$.

Figure 11: $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ and $\mathcal{U}_{\langle \mathcal{T}_2, \mathcal{A}_2 \rangle}$ for $\phi = \exists X_1 \forall X_2 \exists X_3 (X_1 \wedge (X_2 \vee \neg X_3))$.

$1 \leq j \leq m$, and $P_i^{k'}, T_i'$ if its end is labeled with Z_i^k, and the labels of edges $\longrightarrow$ are shown to the left of each infinite and finite path. The labels of the nodes (if any) are shown next to each node.

Since ϕ is valid, it follows that $\langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$. In fact, $\mathcal{U}_{\langle \mathcal{T}_2, \mathcal{A}_2 \rangle}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$. ∎

Next we provide an upper bound for the membership problem for universal UCQ-solutions with simple ABoxes. In the following given KBs $\mathcal{K}_1$ and $\mathcal{K}_2$, and a signature Σ, we are going to develop an algorithm for checking whether $\mathcal{U}_{\mathcal{K}_1}$ is finitely Σ-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_2}$, based on a reduction to infinite games, which extends the reduction in Section 5.2 from the universal models of ABoxes (i.e., the graph of constants) to the universal models of KBs (i.e., possibly infinite forests hanging from the graph of constants). Let $\mathcal{G}_i$ with $\Delta^{\mathcal{G}_i} = \mathsf{Ind}(\mathcal{K}_i) \cup \mathsf{Wit}(\mathcal{K}_i)$ be the generating model and $\mathcal{U}_i$ the canonical model of $\mathcal{K}_i$. Moreover, let $\bar{\mathbf{r}}_\Sigma^{\mathcal{G}_i}(u,v)$ contain the inverses of the roles in $\mathbf{r}_\Sigma^{\mathcal{G}_i}(u,v)$. We begin with a very simple extension of the game $G_\Sigma^c(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$.

INFINITE GAME $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ is a pair (G_i, F_i), where G_i is the *game graph*, and F_i is the set of states that Spoiler wants to reach.

The game graph $\mathsf{G}_i = (\mathsf{S}, \mathsf{D}, \mathsf{T})$ has the set of states of the kind $(u \mapsto \sigma)$ and $(\sigma,\ u \rightsquigarrow u')$, where $u, u' \in \Delta^{\mathcal{G}_1}$ and $\sigma \in \Delta^{\mathcal{U}_2}$.

- S consists of the states $(u \mapsto \sigma)$ with $\mathbf{t}_\Sigma^{\mathcal{G}_1}(u) \subseteq \mathbf{t}_\Sigma^{\mathcal{U}_2}(\sigma)$ and $\sigma = u$ if $u \in \mathsf{Ind}(\mathcal{K}_1)$; intuitively, such states represent a mapping of $\delta \in \Delta^{\mathcal{U}_1}$ with $\mathsf{tail}(\delta) = u$ to σ. Given this partial homomorphism, Spoiler can decide to challenge Duplicator with one of the successors u' of u in $\mathcal{G}_1$.
- D consists of the states $(\sigma,\ u \rightsquigarrow u')$ with $u \rightsquigarrow_{\mathcal{K}_1} u'$ such that $u \xrightarrow[\Sigma]{} u'$; these states represent "challenges" that Duplicator must address by finding in $\mathcal{U}_2$ a neighbor σ' of σ so that the "challenged" edge $(\delta,\ \delta \cdot u')$ of $\mathcal{U}_1$ can be "mapped" to $(\sigma,\ \sigma')$.

Therefore, the transitions between S and D, forming T, are defined as the union of:

- $\big((u \mapsto \sigma),\ (\sigma,\ u \rightsquigarrow u')\big)$, and
- $\big((\sigma,\ u \rightsquigarrow u'),\ (u' \mapsto \sigma')\big)$ whenever $\mathbf{r}_\Sigma^{\mathcal{G}_1}(u,u') \subseteq \mathbf{r}_\Sigma^{\mathcal{U}_2}(\sigma,\sigma')$.

Notice that the size of G is $O(|\mathcal{G}_1| \times |\mathcal{U}_2|)$.

The set F_i, which is the set of states that Spoiler wants to reach, is given by the duplicating states that are "dead ends", i.e.,

$$F_i = \big\{(\sigma,\ u \rightsquigarrow u') \mid (u' \mapsto \sigma') \notin \mathsf{S} \ \text{ or } \ \mathbf{r}_\Sigma^{\mathcal{G}_1}(u,u') \not\subseteq \mathbf{r}_\Sigma^{\mathcal{U}_2}(\sigma,\sigma'), \text{for all } \sigma' \in \Delta^{\mathcal{U}_2}\big\}.$$

The game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ proceeds slightly differently from $G_\Sigma^c(\mathcal{G}_\mathcal{K}, \mathcal{U}_\mathcal{A})$. For each $n \geq 1$ and $u_0 \in \Delta^{\mathcal{G}_1}$, Duplicator tries to construct a Σ-homomorphism from $\mathcal{U}_1^{u_0,n}$ to $\mathcal{U}_2$, where $\mathcal{U}_1^{u_0,n}$ is a subinterpretation of $\mathcal{U}_1$ such that there exists $\delta_0 \in \Delta^{\mathcal{U}_1}$ with $\mathsf{tail}(\delta_0) = u_0$, and $\delta = \delta_0 \cdot \delta'$ and $\mathsf{len}(\delta') \leq n$, for each $\delta \in \Delta^{\mathcal{U}_1^{u_0,n}}$. At the same time, Spoiler attempts to fail him by finding a path in $\mathcal{U}_1$ that does not have a homomorphic image in $\mathcal{U}_2$, given the partial homomorphism constructed so far. Thus, Spoiler starts in a state $\mathfrak{s}_0 = (u_0 \mapsto \sigma_0)$ for some σ_0 with $u_0 = \sigma_0$ if $u_0 \in \mathsf{Ind}(\mathcal{K}_1)$, which corresponds to setting the homomorphic image of u_0 to σ_0, and in each his turn chooses a successor u_{i+1}

of u_i in $\mathcal{G}_1$: the “challenge” represented by the state $(\sigma_i,\, u_i \rightsquigarrow u_{i+1})$. Then Duplicator tries to find $\sigma_{i+1} \in \Delta^{\mathcal{U}_2}$ that could be the image of the “challenged” node $u_0 \cdots u_{i+1}$, which gives the next Spoiler’s state $(u_{i+1} \mapsto \sigma_{i+1})$. Duplicator loses if at some point he cannot find where to map the challenged node, i.e., the game reached a dead-end of Duplicator: $(\sigma_i,\, u_i \rightsquigarrow u_{i+1}) \in F_i$. Otherwise, they can reach a dead-end of Spoiler, or continue until the n-th successor of u_0, hence Duplicator wins.

Therefore, for an ordinal $\lambda \leq \omega$, we say that Duplicator has a *λ-winning strategy from $\mathfrak{s}_0 = (u_0 \mapsto \sigma_0)$ in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$* if, for $i < \lambda$ and every play

$$(u_0 \mapsto \sigma_0), (\sigma_0,\, u_0 \rightsquigarrow u_1), \dots, (u_i \mapsto \sigma_i), (\sigma_i,\, u_i \rightsquigarrow u_{i+1})$$

conform with this strategy, $(\sigma_i,\, u_i \rightsquigarrow u_{i+1})$ is not a dead-end of Duplicator.

We prove that verifying whether $\mathcal{U}_1$ can be finitely Σ-homomorphically mapped to $\mathcal{U}_2$ reduces to checking whether Duplicator has a n-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ from some state $\mathfrak{s}_0^n = (u_0 \mapsto \sigma_0)$.

Lemma 6.2.3. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be KBs with the universal models $\mathcal{U}_1$ and $\mathcal{U}_2$ respectively, $\mathcal{G}_1$ the generating model of $\mathcal{K}_1$, and Σ a signature. Then $\mathcal{U}_1$ is finitely Σ-homomorphically embeddable into $\mathcal{U}_2$ iff*

(ABOX) $\mathbf{r}_\Sigma^{\mathcal{U}_1}(a,b) \subseteq \mathbf{r}_\Sigma^{\mathcal{U}_2}(a,b)$, *for all $a, b \in \mathsf{Ind}(\mathcal{K}_1)$;*

(WIN) *for each $u_0 \in \Delta^{\mathcal{G}_1}$ and $n < \omega$ there exists $\sigma_0 \in \Delta^{\mathcal{U}_2}$ such that Duplicator has an n-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ from $\mathfrak{s}_0^n = (u_0 \mapsto \sigma_0)$.*

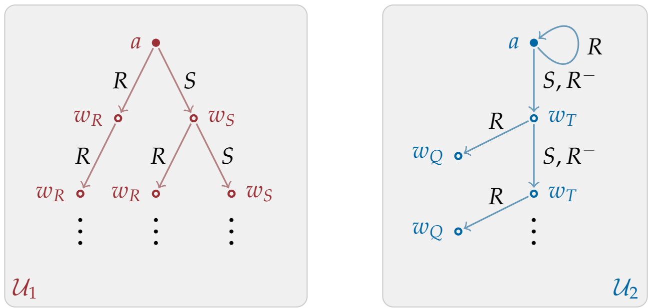
Proof. $(\Rightarrow)$ Suppose $\mathcal{U}_1$ is finitely Σ-homomorphically embeddable into $\mathcal{U}_2$. Then (abox) holds by definition of Σ-homomorphism. To show that (win) holds, suppose $u_0 \in \Delta^{\mathcal{G}_1}$ and $n < \omega$ are given. Take a sub-interpretation $\mathcal{U}_1^{u_0,n}$ of $\mathcal{U}_2$ that contains δu_0, for some (say, the shortest) word δ, and all elements of $\Delta^{\mathcal{U}_1}$ whose distance from δu_0 does not exceed n. Let $h\colon \mathcal{U}_1^{u_0,n} \to \mathcal{U}_2$ be a Σ-homomorphism. Take $\sigma_0 = h(\delta u_0)$: then $(u_0 \mapsto \sigma_0)$ is a state in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$. We show that Duplicator has an n-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ from $(u_0 \mapsto \sigma_0)$. Suppose Spoiler moves to $(\sigma_0,\, u_0 \rightsquigarrow u_1)$. Then $\delta u_0 u_1$ is an element of $\mathcal{U}_1^{u_0,n}$, and Duplicator can respond with $\sigma_1 = h(\delta u_0 u_1)$. Note that $(u_1 \mapsto \sigma_1)$ is a state in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ because h is a Σ-homomorphism. In such a manner, Duplicator can use h to respond to each challenge $(\sigma_i,\, u_i \rightsquigarrow u_{i+1})$ with $0 \leq i \leq n$.

$(\Leftarrow)$ Let $\mathcal{U}_1'$ be a sub-interpretation of $\mathcal{U}_1$ containing n elements or $\mathcal{U}_1$ itself if such a sub-interpretation does not exist. Consider first the case when $\mathcal{U}_1'$ is a tree with the root δu_0 for some $u_0 \in \Delta^{\mathcal{G}_1}$. We define, by induction, a Σ-homomorphism $h\colon \mathcal{U}_1' \to \mathcal{U}_2$ as follows. Take an n-winning strategy for Duplicator in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ starting from a suitable state $(u_0 \mapsto \sigma_0)$ and set $h(\delta u_0) = \sigma_0$. Suppose now that $\delta u_0 \dots u_k$ is an element of $\mathcal{U}_1'$ such that whenever $u_i \xrightarrow[\Sigma]{} \dots \xrightarrow[\Sigma]{} u_j$ for some $0 \leq i \leq j < k$ there is a play $(u_i \mapsto \sigma_i), (\sigma_i,\, u_i \rightsquigarrow u_{i+1}), \dots, (u_j \mapsto \sigma_j)$, which conforms with some n-winning strategy. Assume $u_{k-1} \xrightarrow[\Sigma]{} u_k$, then $(\sigma_{k-1},\, u_{k-1} \rightsquigarrow u_k)$ is a valid challenge for Duplicator in some n-winning strategy from the state $(u_{k-1} \mapsto \sigma_{k-1})$ and consider the reply $(u_k \mapsto \sigma_k)$ of Duplicator: we set $h(\delta u_0 \dots u_k) = \sigma_k$. By construction of $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$, we have that h is a Σ-homomorphism. If however $u_{k-1} \overset{}{\underset{\Sigma}{\dashrightarrow}} u$ for each $u \in \Delta^{\mathcal{G}_1}$ such that $u_{k-1} \rightsquigarrow_{\mathcal{K}_1} u$, then we can set $h(\delta u_0 \dots u_k) = \sigma'$, where $\sigma' \in \Delta^{\mathcal{U}_2}$ is such that there is an n-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ from $(u_k \mapsto \sigma')$ (such σ' exists by (win)).

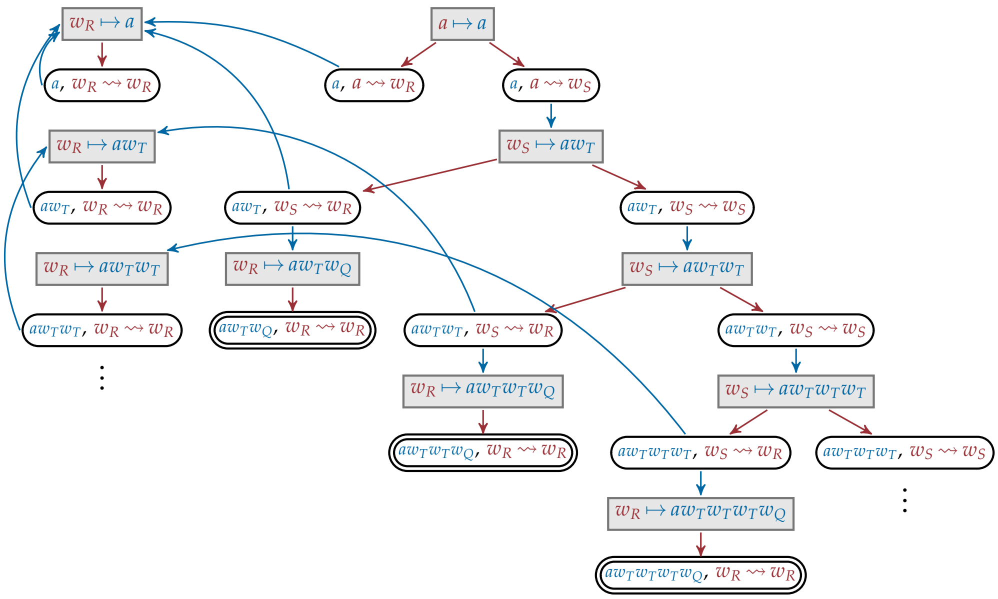
An arbitrary finite sub-interpretation of $\mathcal{U}_1$ can be represented as a union of finitely many maximal trees in which ABox individuals can only be roots. Let h be the union of the corresponding Σ-homomorphisms for all these sub-trees. In view of (abox), h is a Σ-homomorphism. $\square$

The following two examples illustrate the reduction, moreover we discuss the type of homomorphisms and games in each of the cases.

Example 6.2.4. Suppose $\Sigma = \{R, S\}$, $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ and $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$, where $\mathcal{T}_1 = \{\exists R^- \sqsubseteq \exists R,\ \exists S^- \sqsubseteq \exists S,\ \exists S^- \sqsubseteq \exists R\}$, $\mathcal{A}_1 = \{\exists R(a),\ \exists S(a)\}$, and $\mathcal{T}_2 = \{\exists T^- \sqsubseteq \exists T,\ \exists T^- \sqsubseteq \exists Q,\ T \sqsubseteq S,\ T \sqsubseteq R^-,\ Q \sqsubseteq R\}$, $\mathcal{A}_2 = \{R(a,a),\ \exists T(a)\}$. The (projections of the) canonical models $\mathcal{U}_1$ and $\mathcal{U}_2$ look as follows:



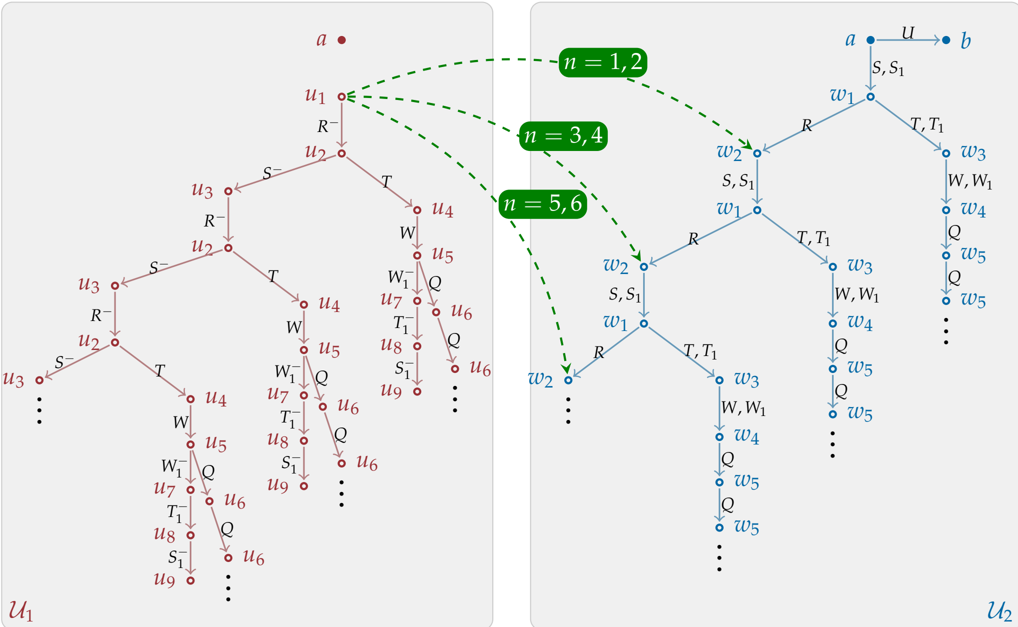
and it is easy to see that $\mathcal{U}_1$ is (finitely) Σ-homomorphically embeddable into $\mathcal{U}_2$. Such a homomorphism "starts" in a, then "goes down" (i.e., forward) the tree $\mathcal{U}_2$ to $aw_T \cdots w_T$, next "goes up" (i.e., backward) to a, and finally, "stays" in a. The corresponding game graph of $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ is shown in the picture below. Here, the moves of Spoiler are shown in red color, and the moves of Duplicator are shown in blue color.



Observe that Duplicator has an n-winning strategy from $(a \mapsto a)$ in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ for each $n \geq 1$. He has no n-winning strategy, however, from $(w_R \mapsto aw_Tw_Q)$ for $n \geq 1$. $\blacksquare$

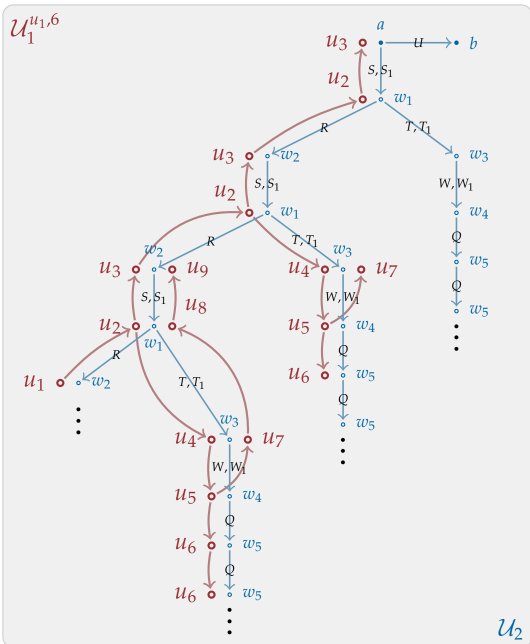
In the example above, the winning Duplicator strategy from $(a \mapsto a)$ follows the pattern of the homomorphism. We call such strategies *start-bounded*: a strategy for Duplicator in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ starting from a state $(u_0 \mapsto \sigma_0)$ is *start-bounded* if it never leads to $(u_i \mapsto \sigma_i)$ such that $\sigma_0 = \sigma_i w$, for some w and $i > 0$. In other words, Duplicator cannot use those elements of $\mathcal{U}_2$ that are located closer to the ABox than σ_0; the ABox individuals in $\mathcal{U}_2$ can only be used if $\sigma_0 \in \mathsf{Ind}(\mathcal{K}_2)$.

Example 6.2.5. Consider the (projections over Σ of the) canonical models $\mathcal{U}_1$ and $\mathcal{U}_2$ depicted as follows:



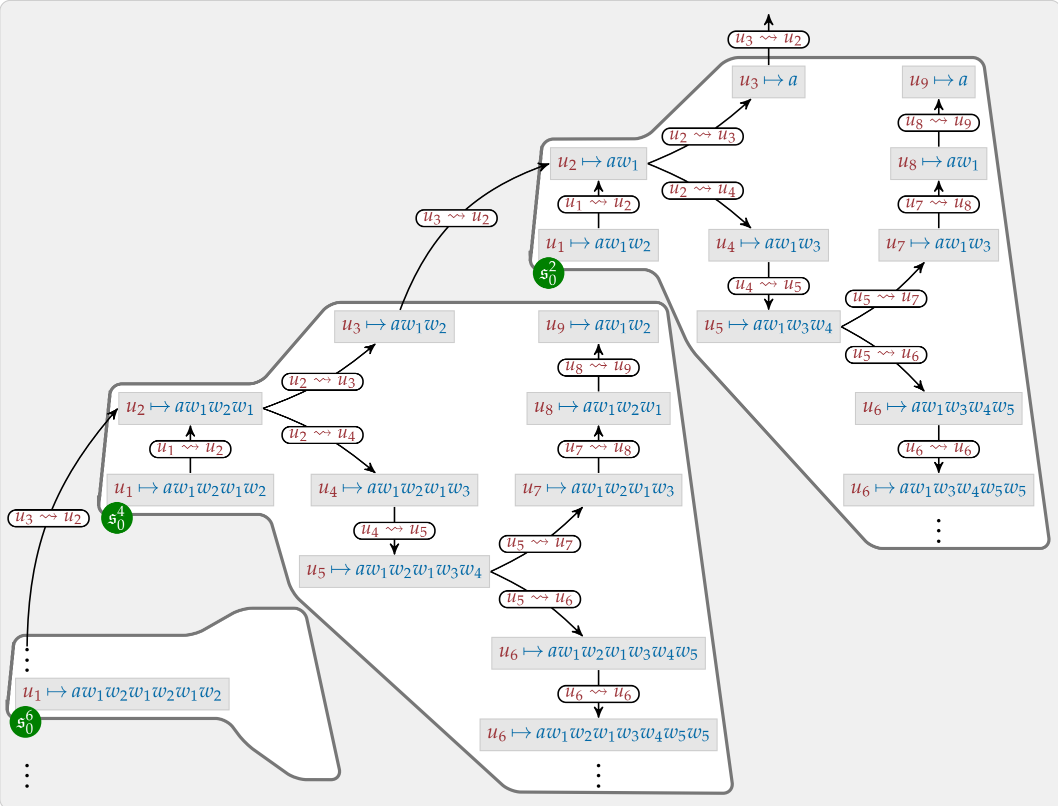
One can see that $\mathcal{U}_1$ is finitely Σ-homomorphically embeddable into $\mathcal{U}_2$. In particular, for $n = 1,2$, a Σ-homomorphism from $\mathcal{U}_1^{u_1,n}$ to $\mathcal{U}_2$ maps au_1 to aw_1w_2; for $n = 3,4$, the element au_1 can be mapped to $aw_1w_2w_1w_2$; for $n = 5,6$, the element au_1 can be mapped to $aw_1w_2w_1w_2w_1w_2$, and so on.

Thus, if we view homomorphism as wrapping a wire ($\mathcal{U}_1$) around a tree ($\mathcal{U}_2$), then a Σ-homomorphism from $\mathcal{U}_1^{u_1,6}$ to $\mathcal{U}_2$ can be depicted as follows:



The homomorphism in this case has a different pattern from the previous example: it starts in $aw_1w_2 \cdots w_1w_2$, then goes up and eventually "reaches" a; on the way it also "forks" to go both down and up. Note that when going up, some elements are mapped to the image of their predecessor ($au_1u_2u_4u_5u_7u_8$), and some elements are mapped to the image of other elements that "started" going up "earlier" ($au_1u_2u_4u_5u_7u_8u_9$).

The corresponding game graph of $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ can be partially depicted as follows.



Here to save space, since the choices for mapping the elements of $\Delta^{\mathcal{U}_1}$ are always unique, we depict the graph slightly differently as usual: the duplicating states are shown as labels on the transitions, and the σ component is omitted. Observe that in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$, Duplicator has an n-winning strategy from $\mathfrak{s}_0^2 = (u_1 \mapsto aw_1w_2)$ for $n = 1, 2$; from $\mathfrak{s}_0^4 = (u_1 \mapsto aw_1w_2w_1w_2)$ for $n = 3, 4$; from $\mathfrak{s}_0^6 = (u_1 \mapsto aw_1w_2w_1w_2w_1w_2)$ for $n = 5, 6$, and so on. ■

One can see that the winning Duplicator strategy from $(u_1 \mapsto aw_1w_2w_1w_2w_1w_2)$ in the example above follows the pattern of the homomorphism as well. This strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ is the most general one, and it is composed of one *backward* strategy and a number of start-bounded strategies. A λ-strategy for Duplicator in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ starting from a state $(u_0 \mapsto \sigma_0)$ is called *backward* if for every play

$$(u_0 \mapsto \sigma_0), (\sigma_0,\ u_0 \rightsquigarrow u_1), \dots, (u_{i-1} \mapsto \sigma_{i-1})$$

with $i - 1 < \lambda$ that conforms with this strategy, and every challenge $(\sigma_{i-1},\ u_{i-1} \rightsquigarrow u_i)$ by Spoiler, the response σ_i of Duplicator is the immediate predecessor of σ_{i-1} in $\mathcal{U}_2$, i.e., $\sigma_{i-1} = \sigma_i w$, for some $w \in \Delta^{\mathcal{G}_1}$ (Duplicator loses in case $\sigma_{i-1} \in \mathsf{Ind}(\mathcal{K}_2)$). Note that, since $\mathcal{U}_2$ is tree-shaped, the response of Duplicator to any other challenge $(\sigma_{i-1},\ u_{i-1} \rightsquigarrow u'_i)$ must be the same σ_i.

Unfortunately, characterization in Lemma 6.2.3 does not provide an algorithm for checking finite Σ-homomorphisms, as in general, the game graph of $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ is infinite. To deal with that, in the following, we are going to define two games "played" on the generating models $\mathcal{G}_1$ and $\mathcal{G}_2$. The main idea of these games is that every element of $\Delta^{\mathcal{U}_2}$ is "visited" only once. First, we define so called *start-bounded* games that capture homomorphisms as in Example 6.2.4 and cover *start-bounded* strategies in $\mathsf{G}_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$. Then, we define *finite* games that capture arbitrary homomorphisms as in Example 6.2.5 and cover arbitrary Duplicator strategies in $\mathsf{G}_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$.

START-BOUNDED GAME $\mathsf{G}^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ is a pair (G, F).

The game graph $\mathsf{G} = (\mathsf{S}, \mathsf{D}, \mathsf{T})$ has the set of states of the kind $(\Gamma, \Xi \mapsto x)$ and $(\Xi, x, u \rightsquigarrow u')$, where $\Xi, \Gamma \subseteq \Delta^{\mathcal{G}_1}$, $\Xi \neq \emptyset$, $u, u' \in \Delta^{\mathcal{G}_1}$, and $x \in \Delta^{\mathcal{G}_2}$.

- S consists of the states $(\Gamma, \Xi \mapsto x)$ with $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(x)$ for each $u \in \Xi$, and $x = u$ if $u \in \Xi \cap \mathsf{Ind}(\mathcal{K}_1)$. Intuitively, Ξ is the set of u that Duplicator "guessed" should be mapped to x. Note that Γ stores previously mapped elements of $\Delta^{\mathcal{G}_1}$, which are used by Spoiler to choose valid challenges.
- D consists of the states $(\Xi, x, u \rightsquigarrow u')$ with $u \in \Xi$, $u \rightsquigarrow_{\mathcal{K}_1} u'$ such that $u \xrightarrow[\Sigma]{} u'$. Intuitively, such states represent challenges that Duplicator has to address. Note that Ξ and x store relevant information used by Duplicator to choose a response.

The transitions in T are defined as:

- $\big((\Gamma, \Xi \mapsto x), (\Xi, x, u \rightsquigarrow u')\big)$ whenever the following condition is satisfied:

  (NBK) if $u' \in \Gamma$ and $x \notin \mathsf{Ind}(\mathcal{K}_2)$, then $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \not\subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(z, x)$ for some $z \in \Delta^{\mathcal{G}_2}$ such that $z \rightsquigarrow_{\mathcal{K}_2} x$.

  In particular, if $u' \in \Gamma$, $x \notin \mathsf{Ind}(\mathcal{K}_2)$, and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(z, x)$, it means that u' has been "already" mapped to the predecessor of x, so it is not a valid challenge.
- for $u' \in \Xi'$ and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \mathbf{r}^{\mathcal{G}_2}_\Sigma(x, x')$,

  $\big((\Xi, x, u \rightsquigarrow u'), (\Xi, \Xi' \mapsto x')\big)$ whenever $x \rightsquigarrow_{\mathcal{K}_2} x'$, and

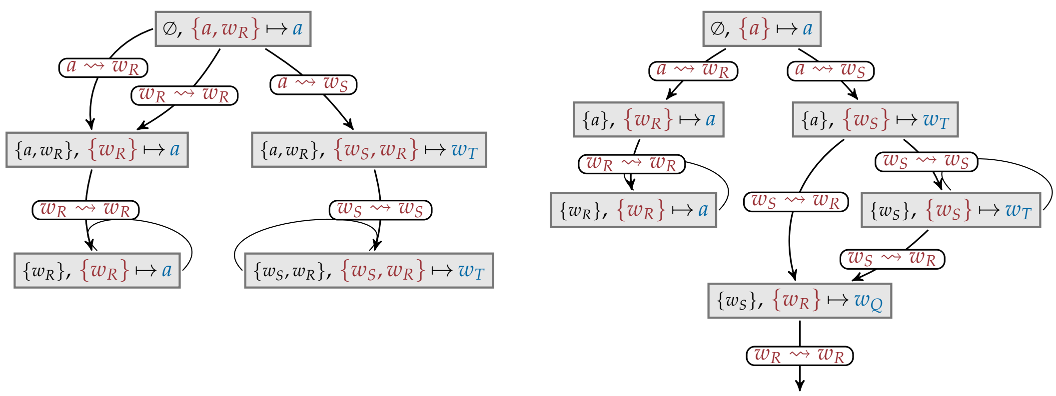
  $\big((\Xi, x, u \rightsquigarrow u'), (\emptyset, \Xi' \mapsto x')\big)$ whenever $x, x' \in \mathsf{Ind}(\mathcal{K}_2)$.

  Observe that because of condition (nbk) Duplicator moves only "forward" in $\mathcal{G}_2$, but has to guess appropriate sets Ξ' in advance.

Notice that the size of G is $2^{O(|\mathcal{G}_1| \times |\mathcal{G}_2|)}$.

As usual, the set F_f consists of the states from D that are "dead ends".

Example 6.2.6. Consider KBs $\mathcal{K}_1$ and $\mathcal{K}_2$ from Example 6.2.4. Then a part of the game graph of $\mathsf{G}^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ can be depicted as in the picture below. Again, the duplicating states are shown as labels on the transitions, with the first two components omitted.



Observe that Duplicator has an ω-winning strategy from $(\emptyset,\ \{a, w_R\} \mapsto a)$ in game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$. He has no ω-winning strategy, however, from $(\emptyset,\ \{a\} \mapsto a)$. Note the crucial guesses $\{a, w_R\} \mapsto a$ and $\{w_S, w_R\} \mapsto w_T$ in the former case. If the game started in $(\emptyset,\ \{a\} \mapsto a)$ or Duplicator responded with $(\{a, w_R\},\ \{w_S\} \mapsto w_T)$ (and failed to guess that w_R must also be mapped to w_T), then after the challenge $w_S \rightsquigarrow w_R$ and response $(\{w_s\},\ \{w_R\} \mapsto w_Q)$, Spoiler would challenge with $w_R \rightsquigarrow w_R$, to which Duplicator could not respond. ■

FINITE GAME $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ is a pair (G_f, F_f).

The game graph $\mathsf{G}_f = (\mathsf{S}, \mathsf{D}, \mathsf{T})$ has the set of states of the kind $(\Xi \mapsto x,\ \Psi)$, and $(x,\ \Xi \rightsquigarrow \Psi)$, where $\Xi, \Psi \subseteq \Delta^{\mathcal{G}_1}$, $x \in \Delta^{\mathcal{G}_2}$, and two special states *loop* and *deadend*.

- S consists of the states $(\Xi \mapsto x,\ \Psi)$ with $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(x)$ for each $u \in \Xi$, $x = u$ if $u \in \Xi \cap \mathsf{Ind}(\mathcal{K}_1)$, and $\Psi \subseteq \Xi^{\rightsquigarrow}$. Note that Ψ contains the challenges that should be mapped to the predecessor of x (i.e., continue going backwards), and consequently, $\Xi^{\rightsquigarrow} \setminus \Psi$ contains the initial challenges in start-bounded games.

- D consists of the states $(x,\ \Xi \rightsquigarrow \Psi)$ with $\Psi \subseteq \Xi^{\rightsquigarrow}$, and the states *loop*, *deadend*,

where $\Xi^{\rightsquigarrow} = \{u' \in \Delta^{\mathcal{G}_1} \mid u \rightsquigarrow_{\mathcal{K}_1} u', \text{ for some } u \in \Xi\}$.

The transitions in T are defined as the union of:

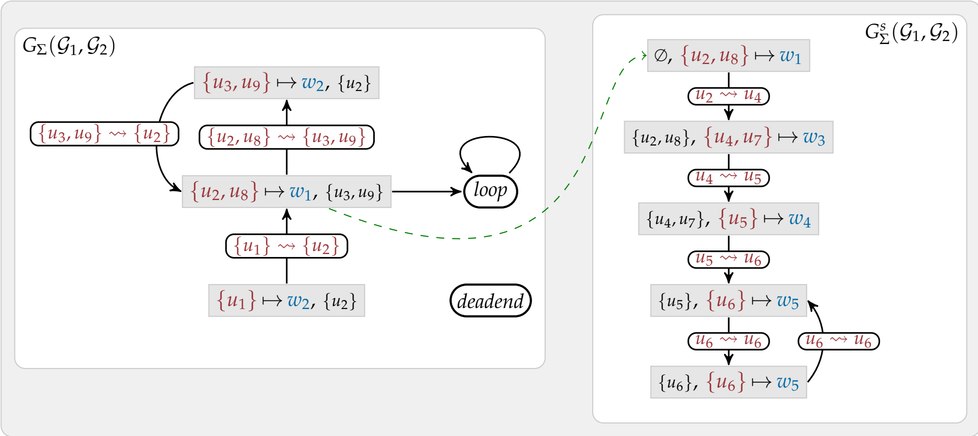
- $\big((\Xi \mapsto x,\ \Psi),\ (x,\ \Xi \rightsquigarrow \Psi)\big)$,
- $\big((x,\ \Xi \rightsquigarrow \Psi),\ (\Xi' \mapsto x',\ \Psi')\big)$ if $\Psi \subseteq \Xi'$, $x' \rightsquigarrow_{\mathcal{K}_2} x$ and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(x', x)$,
- $\big((\Xi \mapsto x,\ \Psi),\ \textit{loop}\big)$ if Duplicator has an ω-winning strategy from $(\emptyset,\ \Xi \mapsto x)$ in the game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ with the initial challenges from $\Xi^{\rightsquigarrow} \setminus \Psi$,
- $\big((\Xi \mapsto x,\ \Psi),\ \textit{deadend}\big)$ if Duplicator has *no* ω-winning strategy from $(\emptyset,\ \Xi \mapsto x)$ in the game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ with the initial challenges from $\Xi^{\rightsquigarrow} \setminus \Psi$, and
- $(\textit{loop},\ \textit{loop})$.

Notice that the size of G_f is $2^{O(|\mathcal{G}_1| \times |\mathcal{G}_2|)}$.

As usual, the set F_f consists of the states from D that are "dead ends".

The game $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ proceeds as follows. For each $u_0 \in \Delta^{\mathcal{G}_1}$, Spoiler starts in a state $\mathfrak{s}_0 = (\Xi_0 \mapsto x_0,\ \Psi_0)$ with $u_0 \in \Xi_0$. In each his turn, if $x_i \in \mathsf{Ind}(\mathcal{K}_2)$, Spoiler "launches" a start-bounded game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ with the initial state $(\emptyset,\ \Xi_i \mapsto x_i)$ and the initial challenges from $\Xi_i^{\rightsquigarrow}$. Otherwise, if $x_i \notin \mathsf{Ind}(\mathcal{K}_2)$, Spoiler chooses whether to challenge the Duplicator with Ψ_i, that is, to continue the game backwards, or to "launch" a start-bounded game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ with the initial state $(\emptyset,\ \Xi_i \mapsto x_i)$ and the initial challenges from $\Xi_i^{\rightsquigarrow} \setminus \Psi_i$. In the former case, Duplicator responds by choosing a predecessor x_{i+1} of x_i and guessing a set Ξ_{i+1}, $\Psi_i \subseteq \Xi_{i+1}$, that should be mapped to x_{i+1}, moreover he guesses Ψ_{i+1}: which successors of of Ξ_{i+1} should continue to be mapped backwards. It gives the next Spoiler's state $(\Xi_{i+1} \mapsto x_{i+1},\ \Psi_{i+1})$. Finite games make "calls" to start-bounded games when it is needed.

Example 6.2.7. Consider $\mathcal{U}_1$ and $\mathcal{U}_2$ from Example 6.2.5. The parts of the game graphs of $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ and $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ that belong to the winning region of Duplicator are shown below.



For $u_1 \in \Delta^{\mathcal{G}_1}$ the finite game starts in the state $(\{u_1\} \mapsto w_2,\ \{u_2\})$, then Duplicator responds to the challenge $\{u_1\} \rightsquigarrow \{u_2\}$ with the state $(\{u_2, u_8\} \mapsto w_1,\ \{u_3, u_9\})$, thus guesses that not only u_2, but also u_8 should be mapped to w_1 and that the successor u_4 of u_2 should be mapped according to the start-bounded strategy. At this point, Spoiler can challenge Duplicator either with $\{u_2, u_8\} \rightsquigarrow \{u_3, u_9\}$ or with $\{u_2, u_8\} \rightsquigarrow \{u_4\}$. In the first case the game continues according to the backward strategy, in the second case, since Duplicator has an ω-winning strategy from $(\emptyset \mapsto \{u_2, u_8\}, w_1)$ in $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ (which can de determined in an external "call"), Spoiler moves to the state *loop* (hence loses). In the game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ from $(\emptyset \mapsto \{u_2, u_8\}, w_1)$ Duplicator does one important guess, namely $\{u_4, u_7\} \mapsto w_3$. ■

Lemma 6.2.8. *For each $u_0 \in \Delta^{\mathcal{G}_2}$ and $n < \omega$ there exists $\sigma_0 \in \Delta^{\mathcal{U}_2}$ such that Duplicator has an (arbitrary) n-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ from $\mathfrak{s}^n_0 = (u_0 \mapsto \sigma_0)$ iff Duplicator has an ω-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ from some state $(\Xi_0 \mapsto x_0,\ \Psi_0)$ with $u_0 \in \Xi_0$.*

Proof. ($\Rightarrow$) Let $\mathbb{S} = \{\mathcal{S}_n \mid n < \omega\}$ be the set of the given n-winning strategies for Duplicator in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$ and suppose that $\mathcal{S}_n$ begins with $(u_0 \mapsto \sigma^n_0)$, $n < \omega$.

We define a (possibly infinite) tree $\mathfrak{T}$ whose nodes are of the form $(u \mapsto z, k)$, where $u \in \Delta^{\mathcal{G}_1}$, z is a suffix of some element in $\Delta^{\mathcal{U}_2}$, $k < \omega$, whose edges are labelled with $u \rightsquigarrow u'$, and the following conditions hold:

(1) the root of $\mathfrak{T}$ is of the form $(u_0 \mapsto w, 0)$, $w \in \Delta^{\mathcal{G}_2}$;

(2) $\mathbf{t}^{\mathcal{G}_1}_\Sigma(u) \subseteq \mathbf{t}^{\mathcal{G}_2}_\Sigma(\mathsf{tail}(z))$;

(3) for each node $(u \mapsto z, k)$ in $\mathfrak{T}$ and each $u \xrightarrow[\Sigma]{}^1 u'$, there is exactly one $(u \rightsquigarrow u')$-successor of $(u \mapsto z, k)$ in $\mathfrak{T}$, which can be of the following forms:
   - $(u' \mapsto w', k+1)$, if $z = w$, $w' \xrightarrow[\Sigma]{}^2 w$, and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(w', w)$;
   - $(u' \mapsto z'w', k)$, if $z = z'w'w$, $w' \xrightarrow[\Sigma]{}^2 w$, and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(w', w)$;
   - $(u' \mapsto zw', k)$, if $z = z'w$, $w \xrightarrow[\Sigma]{}^2 w'$, and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \mathbf{r}^{\mathcal{G}_2}_\Sigma(w, w')$;
   - $(u' \mapsto b, -1)$, if $z = a \in \mathsf{Ind}(\mathcal{K}_2)$, $b \in \mathsf{Ind}(\mathcal{K}_2)$, and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u, u') \subseteq \mathbf{r}^{\mathcal{G}_2}_\Sigma(a, b)$;

(4) for each $k \geq 0$, if $(u \mapsto w, k)$ and $(u' \mapsto w', k)$ are nodes in $\mathfrak{T}$ with $w, w' \in \Delta^{\mathcal{G}_2}$, then $w = w'$.

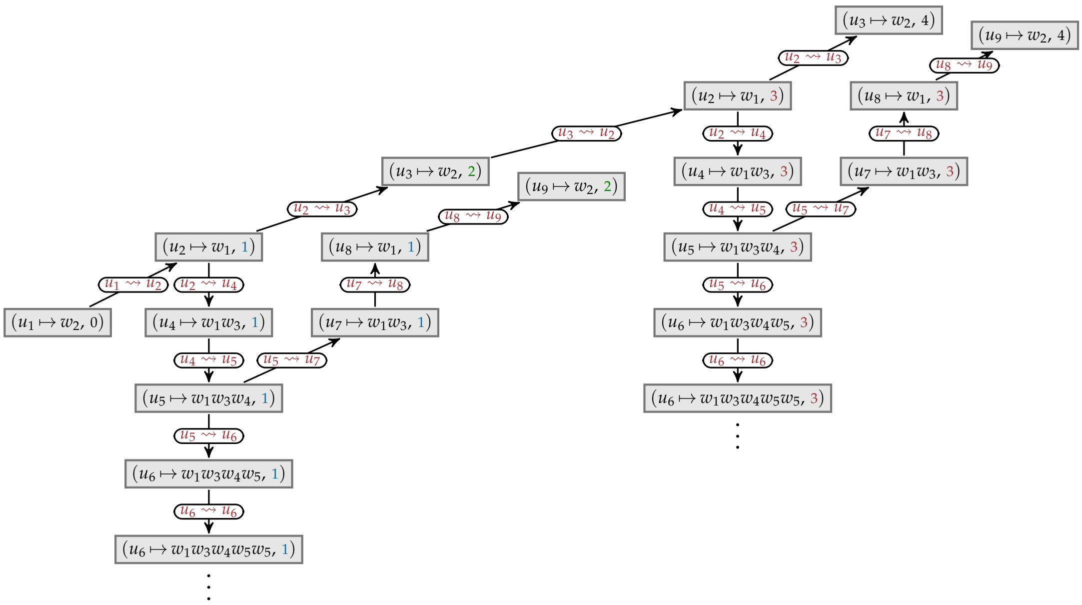


Figure 12: The tree $\mathfrak{T}$ for the game in Example 6.2.5

We call a tree $\mathfrak{T}$ *complete* if whenever a node $(u \mapsto z, k)$ is in $\mathfrak{T}$ and $u \xrightarrow[\Sigma]{}^{1} u'$ then some node $(u' \mapsto z', k')$ is its $(u \rightsquigarrow u')$-successor in $\mathfrak{T}$. Next, for $\mathcal{S} \in \mathbb{S}$, we say that $\mathcal{S}$ *respects* $\mathfrak{T}$ if there exists a map $f_{\mathcal{S}} \colon \{(z,k) \mid (u \mapsto z, k) \in \mathfrak{T}\} \to \Delta^{\mathcal{U}_2}$ such that:

1. $f_{\mathcal{S}}(z,k) = \delta z$, for some δ;
2. $(u \mapsto f_{\mathcal{S}}(z,k))$ is in $\mathcal{S}$, for each $(u \mapsto z, k)$ in $\mathfrak{T}$;
3. if $(u' \mapsto z', k')$ is a $(u \rightsquigarrow u')$-successor of $(u \mapsto z, k)$ in $\mathfrak{T}$, then according to $\mathcal{S}$, Duplicator responds to the challenge $(f_{\mathcal{S}}(z,k),\ u \rightsquigarrow u')$ with $(u' \mapsto f_{\mathcal{S}}(z',k'))$.

It will be shown later that given a complete tree $\mathfrak{T}$ we can construct an ω-winning strategy starting from some $(\Xi_0 \mapsto x_0,\ \Psi_0)$ in the game $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$. But first we show how to construct such a tree $\mathfrak{T}$ using $\mathbb{S}$.

The set $\mathbb{S}$ contains an n-winning strategy starting from $(u_0 \mapsto \sigma_0^n)$, for each $n < \omega$. As $\mathcal{G}_2$ is finite, we can find some x_0 such that $x_0 = \mathsf{tail}(\sigma_0^n)$ for infinitely many n. Denote by $\mathbb{S}_0$ the set of the corresponding strategies from $\mathbb{S}$. As an m-winning strategy is also an l-winning strategy for each $l \leq m$, $\mathbb{S}_0$ contains an n-winning strategy starting from some $(u_0 \mapsto \delta^n x_0)$, for each $n < \omega$. Define $\mathfrak{T}_0$ to be a tree with a single node $(u_0 \mapsto x_0, 0)$. For every $\mathcal{S} \in \mathbb{S}_0$, we set $f_{\mathcal{S}}(x_0, 0) = \delta_{\mathcal{S}} x_0$, where $\delta_{\mathcal{S}}$ is the corresponding δ^n. Thus, all the strategies in $\mathbb{S}_0$ respect $\mathfrak{T}_0$.

Suppose we have already constructed $\mathfrak{T}_i$ and $\mathbb{S}_i$ such that $\mathbb{S}_i$ contains an n-winning strategy for each $n < \omega$, and all of them respect $\mathfrak{T}_i$. If $\mathfrak{T}_i$ is incomplete then it contains a state $(u \mapsto z, k)$ without a $(u \rightsquigarrow u')$-successor, for some $u \xrightarrow[\Sigma]{}^{1} u'$. (We always take such a state that is nearest to the root.) Suppose $f_{\mathcal{S}}(z,k) = \delta_{\mathcal{S}} w$. Consider the responses $(u' \mapsto \sigma^n)$ to the challenge $u \rightsquigarrow u'$ according to the n-winning strategies in $\mathbb{S}_i$, for $n < \omega$. Take some $w' \in \Delta^{\mathcal{G}_2}$ such that $w' = \mathsf{tail}(\sigma^n)$ for infinitely many n. Denote by $\mathbb{S}_{i+1}$ the set of the corresponding strategies from $\mathbb{S}_i$.

Suppose $w' \xrightarrow[\Sigma]{}^{2} w$. If $z = w$ then we add the node $(u' \mapsto w', k+1)$ as a $(u \rightsquigarrow u')$-successor of $(u \mapsto z, k)$ to $\mathfrak{T}_i$, thus obtaining $\mathfrak{T}_{i+1}$. By definition of the canonical model, we also have $\delta_{\mathcal{S}} = \delta'_{\mathcal{S}} w'$, for all $\mathcal{S} \in \mathbb{S}_{i+1}$. We then set $f_{\mathcal{S}}(w', k+1) = \delta_{\mathcal{S}}$.

If $|z| > 1$ then $z = z'w'w$ and $z'w'$ is a suffix of $\delta_{\mathcal{S}}$. In this case, we add the node $(u' \mapsto z'w', k)$ as a $(u \rightsquigarrow u')$-successor of $(u \mapsto z, k)$ to $\mathfrak{T}_i$, thus obtaining $\mathfrak{T}_{i+1}$, and set $f_{\mathcal{S}}(z'w', k) = \delta_{\mathcal{S}}$.

Suppose $w \xrightarrow[\Sigma]{}^2 w'$. In this case, we add $(u' \mapsto zw', k)$ as a $(u \rightsquigarrow u')$-successor of $(u \mapsto z, k)$ to $\mathfrak{T}_i$, thus obtaining $\mathfrak{T}_{i+1}$, and set $f_{\mathcal{S}}(zw', k) = \delta_{\mathcal{S}} ww'$.

Suppose $w, w' \in \mathsf{Ind}(\mathcal{K}_1)$ (hence, $\delta_{\mathcal{S}}$ is empty). In this case, we add $(u' \mapsto w', -1)$ as a $(u \rightsquigarrow u')$-successor of $(u \mapsto z, k)$ to $\mathfrak{T}_i$, obtaining $\mathfrak{T}_{i+1}$, and set $f_{\mathcal{S}}(w', -1) = w'$.

All $\mathcal{S} \in \mathbb{S}_{i+1}$ respect $\mathfrak{T}_{i+1}$, and it is easy to see that $\mathfrak{T}_{i+1}$ satisfies (4).

We proceed in the same way and construct a sequence of trees $\mathfrak{T}_0 \subseteq \mathfrak{T}_1 \subseteq \dots$ until we reach a complete finite tree $\mathfrak{T}_k$; otherwise we take $\mathfrak{T} = \bigcup_{n<\omega} \mathfrak{T}_n$, which is obviously complete. Thus, e.g., for the game in Example 6.2.5, the tree $\mathfrak{T}$ looks as in Figure 12.

Now we show that Duplicator has an ω-winning strategy starting from some state $(\Xi_0 \mapsto x_0,\ \Psi_0)$ in the game $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$. Suppose that we have a complete tree $\mathfrak{T}$ with the root $(u_0 \mapsto x_0, 0)$. We set:

$$
\begin{aligned}
\Xi_0 &= \{u \mid (u \mapsto x_0, 0) \in \mathfrak{T}\},\\
\Phi_0 &= \{u' \mid u \xrightarrow[\Sigma]{1} u',\ u \in \Xi_0,\ (u' \mapsto x_0 w, 0) \in \mathfrak{T}\},\\
&\quad \cup \{u' \mid (u' \mapsto b, -1) \in \mathfrak{T} \text{ is a } (u \rightsquigarrow u')\text{-successor of } (u \mapsto x_0, 0) \in \mathfrak{T}\},\\
\Psi_0 &= \{u' \mid u \xrightarrow[\Sigma]{1} u',\ u \in \Xi_0,\ (u' \mapsto w, 1) \in \mathfrak{T}\}.
\end{aligned}
$$

Note that, by (4), if $(u \mapsto x, 0) \in \mathfrak{T}$ (and $|x| = 1$, that is, $x \in \Delta^{\mathcal{G}_2}$) then $x = x_0$. Moreover, if $x_0 \in \mathsf{Ind}(\mathcal{K}_2)$, then $\Psi_0 = \emptyset$.

Then, for each $i > 0$ such that $\mathfrak{T}$ contains some $(u \mapsto x, i)$, $|x| = 1$, and $x_{i-1} \notin \mathsf{Ind}(\mathcal{K}_2)$, we set

$$
\begin{aligned}
\Xi_i &= \{u \mid (u \mapsto x, i) \in \mathfrak{T}\},\\
\Phi_i &= \{u' \mid u \xrightarrow[\Sigma]{1} u',\ u \in \Xi_i,\ (u' \mapsto xw, i) \in \mathfrak{T}\}\\
&\quad \cup \{u' \mid (u' \mapsto b, -1) \in \mathfrak{T} \text{ is a } (u \rightsquigarrow u')\text{-successor of } (u \mapsto x, i) \in \mathfrak{T}\},\\
\Psi_i &= \{u' \mid u \xrightarrow[\Sigma]{1} u',\ u \in \Xi_i,\ (u' \mapsto w, i+1) \in \mathfrak{T}\}.
\end{aligned}
$$

Note that, by (4), all $(u \mapsto x, i) \in \mathfrak{T}$ with $x \in \Delta^{\mathcal{G}_2}$ share the same x, which we denote by x_i. And again, if $x_i \in \mathsf{Ind}(\mathcal{K}_2)$, then $\Psi_i = \emptyset$.

By (3), the states $\mathfrak{s}_i = (\Xi_i \mapsto x_i,\ \Psi_i)$ define the backward part of an ω-winning strategy for Duplicator in the game $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ starting from $\mathfrak{s}_0$. Thus, it remains to define ω-winning strategies for the start-bounded game $G^{\mathsf{s}}_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ starting from states of the form $(\emptyset,\ \Xi_k \mapsto x_k)$ and first-round challenges $u \rightsquigarrow v$ such that $u \in \Xi_k$ and $v \in \Phi_k$.

Let $k \geq 0$ be such that $\Phi_k \neq \emptyset$. We now transform $\mathfrak{T}$ into a tree $\mathfrak{W}_k$ representing an ω-winning strategy for Duplicator in the game $G^{s}_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ starting from $(\emptyset \mapsto \Xi_k, x_k)$ and first-round challenges $u \rightsquigarrow v$ such that $u \in \Xi_k$ and $v \in \Phi_k$. Thus, $(\emptyset,\ \Xi_k \mapsto x_k)$ is the root of $\mathfrak{W}_k$ *associated with* x_k.

Suppose that we have already defined a node $(\Gamma,\ \Xi \mapsto w)$ associated with a word δw. Let $u \in \Xi$ and $u \xrightarrow[\Sigma]{}^1 v$ be such that the node $(u \mapsto \delta w, k')$ in $\mathfrak{T}$, where k' equals to k or -1, has a $(u \rightsquigarrow v)$-successor of the form $(v \mapsto \delta ww', k')$ (if $(\Gamma \mapsto \Xi, w)$ is the root, we also require that $v \in \Phi_k$). Then we add to $\mathfrak{W}_k$ the node $(\Gamma' \mapsto \Xi', w')$, associated with $\delta w w'$, as a $(u \rightsquigarrow v)$-successor of $(\Gamma \mapsto \Xi, w)$, where

$$\begin{aligned} \Xi' &= \{v' \mid (v' \mapsto \delta w w', k') \in \mathfrak{T}\}, \\ \Gamma' &= \Xi. \end{aligned}$$

If $(\Gamma, \Xi \mapsto a)$ is associated with $a \in \mathsf{Ind}(\mathcal{K}_2)$, and the node $(u \mapsto a, k')$ in $\mathfrak{T}$, with $u \in \Xi$, and k' equal to k or -1, has a $(u \rightsquigarrow v)$-successor of the form $(v \mapsto b, -1)$ with $b \in \mathsf{Ind}(\mathcal{K}_2)$ (note that if $(\Gamma, \Xi \mapsto a)$ is the root, then $\Phi_k = \Xi_k^{\rightsquigarrow}$), then we add to $\mathfrak{W}_k$ the node $(\emptyset, \Xi' \mapsto b)$, associated with b, as a $(u \rightsquigarrow v)$-successor of $(\Gamma, \Xi \mapsto a)$, where

$$\Xi' = \{v' \mid (v' \mapsto b, -1) \in \mathfrak{T}\}.$$

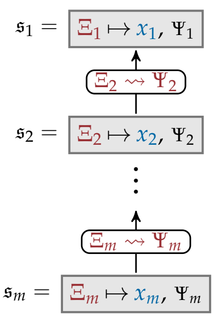
We claim that $\mathfrak{W}_k$ thus constructed represents an ω-winning strategy for Duplicator in the game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ starting from $(\emptyset \mapsto \Xi_k, x_k)$ and first-round challenges $u \rightsquigarrow v$ such that $u \in \Xi_k$ and $v \in \Phi_k$.

($\Leftarrow$) Given $u_0 \in \Delta^{\mathcal{G}_1}$ suppose Duplicator has an ω-winning strategy in $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ from some state $(\Xi_0 \mapsto x_0, \Psi_0)$ such that $u_0 \in \Xi_0$. Let $n < \omega$, we are going to show there is $\sigma_0 \in \Delta^{\mathcal{U}_2}$ such that Duplicator has an n-winning strategy starting from $(u_0 \mapsto \sigma_0)$ in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$.

To define σ_0, consider an N-winning strategy $\mathcal{S}$ of Duplicator from $(\Xi_0 \mapsto x_0, \Psi_0)$ for $N = 2 \times |2^{\mathsf{Wit}(\mathcal{K}_1)}| \times |\mathsf{Wit}(\mathcal{K}_2)| + 1$, and a play

$$(\Xi_m \mapsto x_m, \Psi_m), (x_m, \Xi_m \rightsquigarrow \Psi_m), \dots, (\Xi_2 \mapsto x_2, \Psi_2), (x_2, \Xi_2 \rightsquigarrow \Psi_2), (\Xi_1 \mapsto x_1, \Psi_1)$$

conforming with $\mathcal{S}$ such that $\Xi_m = \Xi_0$, $x_m = x_0$, and $\Psi_m = \Psi_0$. Denote by $\mathfrak{s}_i$ the state $(\Xi_i \mapsto x_i, \Psi_i)$ for $1 \leq i \leq m$.



Then, either $m < N$ and $\Psi_1 = \emptyset$, or $m = N$ and since the number of all possible states in $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ is less than N, there are integers c, r such that $m \geq c > c - r \geq 1$ and $\mathfrak{s}_c = \mathfrak{s}_{c-r}$.

Now, we set $\sigma_0 = \delta'\delta$, where δ' and δ are obtained as follows. In the fist case above, δ is equal to $x_1 \cdots x_m$ and δ' is any (possibly empty) sequence such that $\delta'\delta \in \Delta^{\mathcal{U}_2}$ (such δ' obviously exists). In the second case, δ is equal to the sequence of length $n + 1$:

$$\delta \;=\; x_{c-o}x_{c-o+1} \cdots x_c \;\cdot\; \delta_{c,r} \;\cdots\; \delta_{c,r} \;\cdot\; x_{c+1}x_{c+2} \cdots x_m$$

where $o = (n - (m - c)) \bmod r$, $\delta_{c,r} = x_{c-r+1}x_{c-r+2} \cdots x_c$, and δ' is obtained as before.

Let k be the length of δ, and y_i denote the i-th element of the sequence δ, $1 \leq i \leq k$: $\sigma_0 = \delta' y_1 \cdots y_k$. We define $\mu(i) \in \{1, \dots, m\}$ to be the number such that $y_i = x_{\mu(i)}$. In the first case above $\mu(i) = i$, whereas in the second case

$$\delta = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\multicolumn{1}{c}{1} & \multicolumn{1}{c}{2} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{o+1} & \multicolumn{1}{c}{o+2} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{o+r+1} & \multicolumn{1}{c}{o+r+2} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{o+2r+1} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{n-m+c+2} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{n+1} \\
\hline
x_{c-o} & x_{c-o-1} & \cdots & x_c & x_{c-r+1} & \cdots & x_c & x_{c-r+1} & \cdots & x_c & \multicolumn{1}{c|}{\cdots} & x_{c+1} & \cdots & x_m \\
\hline
\end{array}$$

$$\mu(i) = \begin{cases} c - ((o - i + 1) \bmod r), & \text{for } 1 \leq i \leq n - (m - c) + 1, \\ c + i - 1 - (n - (m - c)), & \text{for } n - (m - c) + 2 \leq i \leq n + 1. \end{cases}$$

Finally, it remains to produce an n-winning strategy $\mathcal{S}'$ of Duplicator from $(u_0 \mapsto \sigma_0)$ in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$. For each challenge of the form $(\sigma_i,\, u_i \rightsquigarrow u_{i+1})$, we are going to define $\sigma_{i+1} \in \Delta^{\mathcal{U}_2}$ so that to set $\mathcal{S}'((\sigma_i,\, u_i \rightsquigarrow u_{i+1})) = (u_{i+1} \mapsto \sigma_{i+1})$. We will also define auxiliary f-values for the Spoiler states $(u_i \mapsto \sigma_i)$ that relate them with the "original" Spoiler states in $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ and $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$.

We first set $f(u_0 \mapsto \sigma_0) = (\Xi_{\mu(k)} \mapsto x_{\mu(k)},\, \Psi_{\mu(k)})$ and consider the challenge $u_0 \rightsquigarrow u_1$ by Spoiler in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$. If $u_1 \in \Psi_{\mu(k)}$, then $k > 1$. We set $\sigma_1 = \delta' y_1 \cdots y_{k-1}$, and $f(u_1 \mapsto \sigma_1) = (\Xi_{\mu(k-1)} \mapsto x_{\mu(k-1)},\, \Psi_{\mu(k-1)})$. If $u_1 \notin \Psi_{\mu(k)}$, then consider the start-bounded game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ with the initial state $(\emptyset,\, \Xi_{\mu(k)} \mapsto x_{\mu(k)})$ and the first-round challenge $u_0 \rightsquigarrow u_1$. Let $(\Gamma,\, \Xi \mapsto z)$ be the response of Duplicator according to $\mathcal{S}$, for some $z \in \Delta^{\mathcal{G}_2}$, $u_1 \in \Xi$. If $z \in \mathsf{Ind}(\mathcal{K}_2)$, we set $\sigma_1 = z$ (note, in this case $\sigma_0 \in \mathsf{Ind}(\mathcal{K}_2)$), otherwise we set $\sigma_1 = \sigma_0 z$. The f-value is defined as $f(u_1 \mapsto \sigma_1) = (\Gamma,\, \Xi \mapsto z)$.

Suppose we defined $\mathcal{S}'((\sigma_{h-1},\, u_{h-1} \rightsquigarrow u_h))$ as $(u_h \mapsto \sigma_h)$, for $h < n$, and the value of f for $(u_h \mapsto \sigma_h)$, moreover assume $\sigma_h = \delta' y_1 \cdots y_{k'} z_1 \cdots z_l$ for $0 \leq k' \leq k$, $l \geq 0$. If now there is no valid challenge $u_h \rightsquigarrow u_{h+1}$, then further moves of Duplicator do not need to be defined. Otherwise consider the challenge $(\sigma_h,\, u_h \rightsquigarrow u_{h+1})$ for Duplicator in $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$.

Assume $f(u_h \mapsto \sigma_h) = (\Gamma',\, \Xi' \mapsto x')$, where $x' = \mathsf{tail}(\sigma_h)$, and by induction hypothesis $u_h \in \Xi'$. Note that in this case, $l \geq 1$. If $u_h \rightsquigarrow u_{h+1}$ is a challenge from $(\Gamma',\, \Xi' \mapsto x')$ in $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$, consider the response $(\Gamma,\, \Xi \mapsto z)$ of Duplicator according to $\mathcal{S}$: if $z \in \mathsf{Ind}(\mathcal{K}_2)$ we set $\sigma_{h+1} = z$, otherwise we set $\sigma_{h+1} = \sigma_h z$. The f-value is defined as $f(u_{h+1} \mapsto \sigma_{h+1}) = (\Gamma,\, \Xi \mapsto z)$. If $u_h \rightsquigarrow u_{h+1}$ is *not* a valid challenge from $(\Gamma',\, \Xi' \mapsto x')$ in $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$, it is the case (nbk) does not hold for $(\Gamma',\, \Xi' \mapsto x')$ and $(\Xi', x',\, u_h \rightsquigarrow u_{h+1})$, which means $x' \notin \mathsf{Ind}(\mathcal{K}_2)$, $u_{h+1} \in \Gamma'$, and $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u_h, u_{h+1}) \subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(z, x')$, where z is the element preceding z_l in σ_h. Two cases are possible:

– $l = 1$, therefore the predecessor of $(\Gamma',\, \Xi' \mapsto x')$ according to $\mathcal{S}$ is the starting state $(\emptyset,\, \Xi_{\mu(k')} \mapsto x_{\mu(k')})$ of the game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$, which has been launched from $(\Xi_{\mu(k')} \mapsto x_{\mu(k')},\, \Psi_{\mu(k')})$ in $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$. It follows, $\mathbf{r}^{\mathcal{G}_1}_\Sigma(u_h, u_{h+1}) \subseteq \bar{\mathbf{r}}^{\mathcal{G}_2}_\Sigma(y_{k'}, z_1)$ and, as $\Gamma' = \Xi_{\mu(k')}$, we have $u_{h+1} \in \Xi_{\mu(k')}$. So we set $\sigma_{h+1} = \delta' y_1 \cdots y_{k'}$, and $f(u_{h+1} \mapsto \sigma_{h+1}) = (\Xi_{\mu(k')} \mapsto x_{\mu(k')},\, \Psi_{\mu(k')})$.

– $l > 1$, we consider the predecessor $(\Gamma,\, \Xi \mapsto x)$ of $(\Gamma',\, \Xi' \mapsto x')$ in $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ according to $\mathcal{S}$, with $x = z_{l-1}$. We have $\Gamma' = \Xi$, hence $u_{h+1} \in \Xi$, so we set $\sigma_{h+1} = \delta' y_1 \cdots y_{k'} z_1 \cdots z_{l-1}$, and $f(u_{h+1} \mapsto \sigma_{h+1}) = (\Gamma,\, \Xi \mapsto x)$.

Alternatively, assume $f(u_h \mapsto \sigma_h) = (\Xi' \mapsto x',\, \Psi')$, where $x' = \mathsf{tail}(\sigma_h)$, and by induction hypothesis $u_h \in \Xi'$. Then $l = 0$ and $(\Xi' \mapsto x',\, \Psi') = (\Xi_{\mu(k')} \mapsto x_{\mu(k')},\, \Psi_{\mu(k')})$. We proceed here as in the base case. If $u_{h+1} \in \Psi_{\mu(k')}$, then $k' > 1$: indeed, by construction of δ, if $k = m$, then $\Psi_1 = \emptyset$, otherwise $k = n + 1$, so provided that $h \leq n$, it cannot be the case $k' = 1$. We set $\sigma_{h+1} = \delta' y_1 \cdots y_{k'-1}$, and $f(u_{h+1} \mapsto \sigma_{h+1}) = (\Xi_{\mu(k'-1)} \mapsto x_{\mu(k'-1)}, \Psi_{\mu(k'-1)})$. If $u_{h+1} \notin \Psi_{\mu(k')}$, then consider the start-bounded game $G^s_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$ with the initial state $(\emptyset, \Xi_{\mu(k')} \mapsto x_{\mu(k')})$ and the first-round challenge $u_h \rightsquigarrow u_{h+1}$. Let $(\Gamma, \Xi \mapsto z)$ be the response of Duplicator according to $\mathcal{S}$, for some $z \in \Delta^{\mathcal{G}_2}$, $u_{h+1} \in \Xi$. If $z \in \mathsf{Ind}(\mathcal{K}_2)$, we set $\sigma_{h+1} = z$, otherwise we set $\sigma_{h+1} = \sigma_h z$. The f-value is defined as $f(u_{h+1} \mapsto \sigma_{h+1}) = (\Gamma, \Xi \mapsto z)$.

We have constructed the strategy $\mathcal{S}'$ from $(u_0 \mapsto \sigma_0)$ in the game $G_\Sigma(\mathcal{G}_1, \mathcal{U}_2)$. It can be straightforwardly verified that $\mathcal{S}'$ is n-winning. □

The latter condition in the lemma above can be checked in time $O(|\mathsf{Ind}(\mathcal{K}_1)| \times 2^{|\Delta^{\mathcal{G}_1} \setminus \mathsf{Ind}(\mathcal{K}_1)|} \times |\Delta^{\mathcal{G}_2}|)$, which can be readily seen by analysing the game graph for $G_\Sigma(\mathcal{G}_1, \mathcal{G}_2)$. Therefore we obtain the following upper bound.

Theorem 6.2.9. *The membership problem for universal* UCQ*-solutions with simple ABoxes is in* EXPTIME.

DISCUSSION ON THE TECHNIQUES. We would like to comment on various EXPTIME techniques that are capable of traversing infinite trees, both in forward and backward directions. Those are μ-calculus [80], Guarded Fixed Point logic [63], two-way alternating automata (employed in Section 5.3.2) [102], pushdown processes [105, 93], and infinite two player games on directed (of exponential size) graphs [73, 87]. Thus, for instance, the automata in Section 5.3 could be easily adapted to check general homomorphisms (however there is no elegant way to check finite homomorphisms), while start-bounded games can be described using pushdown processes that use a stack to remember the current path in $\mathcal{U}_2$. We chose the games as they are easy to understand and provide the necessary means to encode the desired behavior (by changing the definition of the game graph).

6.2.2 *The non-emptiness problem*

In this section we show that the non-emptiness problem for universal UCQ-solutions with simple ABoxes is solvable in exponential time. To do so, we exploit the algorithm devised in the previous section.

Given that we can check in EXPTIME the membership problem for universal UCQ-solutions with simple ABoxes, and each target KB that is a candidate for being a universal UCQ-solution is of polynomial size, we obtain a naïve EXPTIME algorithm for the non-emptiness problem for universal UCQ-solutions with simple ABoxes, which provides an upper bound for the non-emptiness problem.

Theorem 6.2.10. *The non-emptiness problem for universal* UCQ*-solutions with simple ABoxes is in* EXPTIME.

Proof. Assume given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a KB $\mathcal{K}_1$ over Σ_1. Then if $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is a universal UCQ-solution with simple ABoxes for $\mathcal{K}_1$ under $\mathcal{M}$, it must be the case that $\mathsf{Ind}(\mathcal{A}_2) \subseteq \mathsf{Ind}(\mathcal{A}_1)$, hence $\mathcal{A}_2$ is of polynomial size. Next, $\mathcal{T}_2$ is a *DL-Lite*$_\mathcal{R}$ TBox defined over the signature Σ_2, so $\mathcal{T}_2$ must be of polynomial size as well. Now, a naïve EXPTIME algorithm for checking the non-emptiness problem, first, guesses a target KB $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ with $\mathsf{Ind}(\mathcal{A}_2) \subseteq \mathsf{Ind}(\mathcal{A}_1)$ (in NP), then checks whether $\mathcal{K}_2$ is a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$ (in EXPTIME). □

The exact complexity of this problem could, however, turn out to be simpler, and it is left as an open problem.

COMPUTING UNIVERSAL UCQ-SOLUTIONS WITH SIMPLE ABOXES. We present an algorithm for computing a universal UCQ-solution for a given source KB $\mathcal{K}_1$ and a mapping $\mathcal{M}$, which can be extracted from the proof of Theorem 6.2.10. The algorithm is presented in Figure 14.

Algorithm ComputeUniversalUCQSolution$(\mathcal{M}, \mathcal{K}_1)$
Input: mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ over Σ_1
Output: a universal UCQ-solution $\mathcal{K}_2$ for $\mathcal{K}_1$ under $\mathcal{M}$ if it exists, **nothing** otherwise.

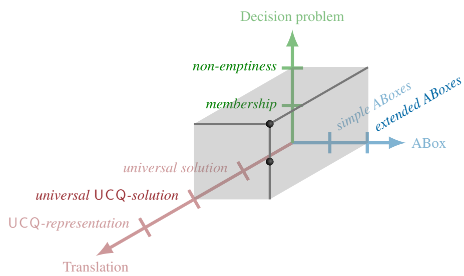
1. Guess an ABox $\mathcal{A}_2$ over Σ_2 such that $\mathsf{Ind}(\mathcal{A}_2) \subseteq \mathsf{Ind}(\mathcal{A}_1)$.
2. Guess a TBox $\mathcal{T}_2$ over Σ_2.

if $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is a universal UCQ-solution for $\mathcal{K}_1$ under $\mathcal{M}$, return $\mathcal{K}_2$.
else return **nothing**.

Figure 13: Algorithm ComputeUniversalUCQSolution.

6.3 UNIVERSAL UCQ-SOLUTIONS WITH EXTENDED ABOXES

In this section, we report on the known results for the membership and non-emptiness problems for universal UCQ-solutions with extended ABoxes. The results for the former problem are carried over from the corresponding problem with simple ABoxes. And as for the latter problem, we provide only a PSpace-lower bound.



6.3.1 *The membership problem*

We observe that the ExpTime algorithm developed in Section 6.2.1 also solves the membership problem for universal UCQ-solutions with extended ABoxes. In fact, assume $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ is a simple KB, $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ is an extended KB, and Σ a signature. Let $\mathcal{U}_1$ and $\mathcal{U}_2$ be the universal models of $\mathcal{K}_1$ and $\mathcal{K}_2$, respectively. Then, a Σ-homomorphism h from $\mathcal{U}_2$ to $\mathcal{U}_1$ is a function from $\Delta^{\mathcal{U}_2}$ to $\Delta^{\mathcal{U}_1}$ such that *(i)* $h(a) = a$ for each $a \in \mathsf{Ind}(\mathcal{K}_2)$, *(ii)* $\mathbf{t}_\Sigma^{\mathcal{U}_2}(x) \subseteq \mathbf{t}_\Sigma^{\mathcal{U}_1}(h(x))$ and $\mathbf{r}_\Sigma^{\mathcal{U}_2}(x,y) \subseteq \mathbf{t}_\Sigma^{\mathcal{U}_1}(h(x), h(y))$ for all $x, y \in \Delta^{\mathcal{U}_2}$. Observe that for $x \in \mathsf{Null}(\mathcal{A}_2)$ there are no extra constraints on $h(x)$, that is, $h(x)$ can be an arbitrary element in $\Delta^{\mathcal{U}_1}$. As the algorithm based on reachability games in Section 6.2.1 "checks" the existence of exactly such homomorphisms, we conclude that this algorithm can be used to solve the membership problem for universal UCQ-solutions with extended ABoxes. Therefore, the following result holds.

Corollary 6.3.1. *The membership problem for universal* UCQ*-solutions with extended ABoxes is in* EXPTIME.

6.3.2 *The non-emptiness problem*

In this section, we prove a PSPACE-lower bound for the non-emptiness problem for universal UCQ-solutions when extended ABoxes are allowed in the target. We do it by reduction from the non-emptiness problem for universal solutions with extended ABoxes, which has been shown to be PSPACE-hard in Section 5.3.2. The main idea of this reduction is that in some cases, even for universal UCQ-solutions, the target TBox must be trivial. So there exists a universal UCQ-solution if and only if there exists a target ABox, finitely Σ_2-homomorphically equivalent to the canonical model of the source KB and the mapping.

Theorem 6.3.2. *The non-emptiness problem for universal* UCQ*-solutions with extended ABoxes is* PSPACE*-hard.*

Proof. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ be a mapping, and $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ a KB over Σ_1. We construct $\mathcal{K}_1'$ and $\mathcal{M}'$ such that there exists a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$ iff there exists a universal UCQ-solution for $\mathcal{K}_1'$ under $\mathcal{M}'$.

Define $\mathcal{M}'$ to be equal to $(\Sigma_1', \Sigma_2', \mathcal{T}_{12}')$, where Σ_1' extends Σ_1 with fresh concept and roles names $\{X_1 \mid X \in \Sigma_2\}$ and fresh role names Q_1, Q_2, Σ_2' extends Σ_2 with a fresh role name Q, and $\mathcal{T}_{12}' = \mathcal{T}_{12} \cup \{X_1 \sqsubseteq X \mid X \in \Sigma_2\} \cup \{Q_1 \sqsubseteq Q, Q_2 \sqsubseteq Q\}$. Let $\mathcal{K}_1' = \langle \mathcal{T}_1', \mathcal{A}_1' \rangle$, where $\mathcal{A}_1'$ is the union of $\mathcal{A}_1$, assertions

$$\{X_1(a_X) \mid X \in \Sigma_2 \text{ is a concept name}\} \cup \{X_1(a_X, b_X) \mid X \in \Sigma_2 \text{ is a role name}\},$$

for fresh constants a_X, b_X for each symbol X, and assertions $\{\exists Q_1(a_Q), Q_2(a_Q, b_Q)\}$, for fresh constants a_Q, b_Q. If $\mathcal{K}_1$ is not Σ_2-safe with respect to $\mathcal{M}$, then $\mathcal{T}_1' = \mathcal{T}_1 \cup \{\exists Q_1^- \sqsubseteq \exists Q_1\}$, otherwise $\mathcal{T}_1' = \mathcal{T}_1$. We prove $\mathcal{K}_1'$ and $\mathcal{M}'$ are as required.

Assume $\mathcal{K}_1$ and $\mathcal{M}$ are inconsistent, that is, the KB $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is inconsistent. Then each inconsistent target KB is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. On the other hand, $\mathcal{K}_1'$ and $\mathcal{M}'$ are inconsistent, and, again, each inconsistent target KB is a universal UCQ-solution for $\mathcal{K}_1'$ under $\mathcal{M}'$. In what follows, we assume $\mathcal{K}_1$ and $\mathcal{M}$ are consistent, and $\mathcal{K}_1'$ and $\mathcal{M}'$ are consistent.

Assume there exists a universal solution $\mathcal{A}_2$ for $\mathcal{K}_1$ under $\mathcal{M}$. Then $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, and it is easy to see that $\mathcal{A}_2 \cup \{X(a_X) \mid X \in \Sigma_2 \text{ is a concept name}\} \cup \{X(a_X, b_X) \mid X \in \Sigma_2 \text{ is a role name}\} \cup \{Q(a_Q, b_Q)\}$ is a universal UCQ-solution for $\mathcal{K}_1'$ under $\mathcal{M}'$.

Now, assume there exists a universal UCQ-solution $\mathcal{K}_2 = \langle \mathcal{T}_2, \mathcal{A}_2 \rangle$ for $\mathcal{K}_1'$ under $\mathcal{M}'$. First, it follows that $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$ does not contain an infinite Q-chain starting from a_Q, hence $\mathcal{T}_1'$ does not contain the axiom $\exists Q_1^- \sqsubseteq \exists Q_1$ and $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$. Second, without loss of generality, we may assume that $\mathcal{T}_2$ does not contain disjointness assertions. Finally, $\mathcal{U}_{\mathcal{K}_2}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$, so for each concept name $A \in \Sigma_2$, $A(a_A) \in \mathcal{A}_2$ and for each role name $P \in \Sigma_2$, $P(a_P, b_P) \in \mathcal{A}_2$. We show that $\mathcal{T}_2$ is a trivial TBox. By contradiction, assume $\alpha \in \mathcal{T}_2$ is a non-trivial axiom. Consider various cases of α:

$\alpha = A \sqsubseteq B$, for concept name B distinct from concept name A. Then $\mathcal{K}_2 \models B(a_A)$, however $\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle \not\models B(a_A)$, hence it is not the case $\mathcal{U}_{\mathcal{K}_2}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Contradiction.

$\alpha = \exists P \sqsubseteq A$, for role name P. Then $\mathcal{K}_2 \models A(a_P)$, however $\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle \not\models A(a_P)$, hence it is not the case $\mathcal{U}_{\mathcal{K}_2}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Contradiction.

$\alpha = \exists P^- \sqsubseteq A$, for role name P. As above, but in this case $\mathcal{K}_2 \models A(b_P)$ and $\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle \not\models A(b_P)$.

$\alpha = P \sqsubseteq R$, for role R distinct from role name P. Then $\mathcal{K}_2 \models R(a_P, b_P)$, however $\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle \not\models R(a_P, b_P)$, hence it is not the case $\mathcal{U}_{\mathcal{K}_2}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Contradiction.

$\alpha = A \sqsubseteq \exists R$, for role R. Then there exists $\sigma \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ distinct from a_A such that $R \in \mathbf{r}^{\mathcal{U}_{\mathcal{K}_2}}(a_A, \sigma)$. Since in $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$, a_A is not connected to anything, $\mathcal{U}_{\mathcal{K}_2}$ is not finitely Σ_2-homomorphically embeddable into $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Contradiction.

$\alpha = \exists P \sqsubseteq \exists R$, for role R distinct from role name P. Then there exists $\sigma \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ distinct from a_P such that $R \in \mathbf{r}^{\mathcal{U}_{\mathcal{K}_2}}(a_P, \sigma)$. If $\sigma = b_P$ then we get a contradiction similar to the case $\alpha = P \sqsubseteq R$. If $\sigma \neq b_P$ then we get a contradiction as above.

$\alpha = \exists P^- \sqsubseteq \exists R$, for role R distinct from P^-. As above.

$\alpha = \exists P^- \sqsubseteq \exists P$, for role name P. Then in $\mathcal{U}_{\mathcal{K}_2}$ there exists an infinite P-chain starting from b_P, and obviously, it is not finitely Σ_2-homomorphically embeddable into $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Contradiction.

Therefore, $\mathcal{T}_2$ is a trivial TBox, so we obtain that $\mathcal{U}_{\mathcal{A}_2}$ is finitely Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Since $\mathcal{U}_{\mathcal{A}_2}$ is finite, it follows $\mathcal{U}_{\mathcal{A}_2}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1' \cup \mathcal{T}_{12}', \mathcal{A}_1' \rangle}$. Let $\mathcal{A}_2^-$ be the subset of $\mathcal{A}_2$ such that $\mathsf{Ind}(\mathcal{A}_2') = \mathsf{Ind}(\mathcal{A}_1)$. It is easy to see that $\mathcal{U}_{\mathcal{A}_2^-}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$, and as $\mathcal{K}_1$ is Σ_2-safe with respect to $\mathcal{M}$, we conclude that $\mathcal{A}_2^-$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. □

Unlike for universal UCQ-solutions with simple ABoxes, for universal UCQ-solutions with extended ABoxes it is not possible to derive a straightforward algorithm for the non-emptiness problem given the algorithm for the membership problem. The non-emptiness problem for universal UCQ-solutions with extended ABoxes is inherently more difficult as there is no bound on the size of the target extended ABox known apriori. Moreover, at the moment we do not have a characterization of the cases when a universal UCQ-solution with extended ABoxes exists.

6.4 UNIVERSAL UCQ-SOLUTIONS IN DL-LITE$_{RDFS}$

Recall that every universal solution is also a universal UCQ-solution. Therefore, as a corollary of Theorem 5.4.1 and Proposition 4.2.1 we get a trivial complexity bound for the non-emptiness problem for universal UCQ-solutions in *DL-Lite*$_{RDFS}$.

Theorem 6.4.1. *The non-emptiness problem for universal solutions in DL-Lite*$_{RDFS}$ *is in* TRIVIAL.

Let us turn now to the membership problem. By Lemma 5.1.2 and by the way homomorphisms are defined on constants, it follows that a simple target ABox $\mathcal{A}_2$ is a universal solution for a source KB $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ if and only if $\mathcal{U}_{\mathcal{A}_2}$ agrees with $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ on concepts and roles from Σ_2. The latter can be checked very efficiently, so we obtain the following complexity bound.

Theorem 6.4.2. *In DL-Lite$_{RDFS}$, the membership problem for universal solutions with simple ABoxes is* NLOGSPACE-*complete.*

Proof. We show the lower bound by reduction from the reachability problem in directed graphs. □

As for the membership problem with extended ABoxes, clearly the upper bound is inherited from the same problem in *DL-Lite$_{\mathcal{R}}$*. To see that the lower bound applies as well, observe that in the reduction from the 3-colorability problem to show that the membership problem with extended ABoxes is NP-hard we use only role inclusions, hence the instance of KB exchange problem we construct is in *DL-Lite$_{RDFS}$*. The theorem below is a straightforward corollary of Theorem 5.3.3.

Theorem 6.4.3. *In DL-Lite$_{RDFS}$, the membership problem for universal solutions with extended ABoxes is* NP-*complete.*

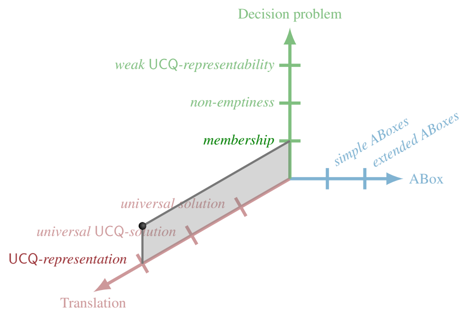
7

REASONING ABOUT UCQ-REPRESENTATIONS

In this chapter we develop the techniques and complexity results for the problem of UCQ-representability. In Sections 7.1 and 7.2, we tackle the membership and non-emptiness problems, respectively, and in Section 7.3 we conclude with the weak UCQ-representability problem.

7.1 THE MEMBERSHIP PROBLEM

In this section we show that the membership problem for UCQ-representations is NLOGSPACE-complete by developing graph-theoretic techniques.



We start by considering the membership problem for UCQ-representations.

One can immediately notice some similarities between this task and the membership problem for universal UCQ-solutions, which was shown to be PSPACE-hard in Theorem 6.2.1. However, the universal quantification over ABoxes in the definition of the notion of UCQ-representation makes the latter problem computationally simpler (in NLOGSPACE instead PSPACE-hard). We now list several observations that help to understand our characterization of UCQ-representations in Lemma 7.1.1, and also to understand why UCQ-representability is a considerably simpler problem than the problem of universal UCQ-solutions.

In the following, assume fixed a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, a source TBox $\mathcal{T}_1$ and a target TBox $\mathcal{T}_2$. We will try to understand what conditions $\mathcal{T}_2$ must satisfy in order to be a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

1) For simplicity, assume here that $\mathcal{T}_1$, $\mathcal{T}_{12}$, and $\mathcal{T}_2$ do not contain disjointness assertions.

Let $\mathcal{A}_1$ be a source singleton ABox, $\mathcal{A}_1 = \{A(a)\}$ for atomic concept A, and assume that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$ for some basic concept B' over Σ_2. Then $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models B'(a)$, and $q = B'(a)$ evaluates to true over $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, i.e., $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models q$. Hence, for $\mathcal{T}_2$ to be a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, it should be the case $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle \models q$. Therefore, from Lemma 6.1.1 it follows that $\mathcal{U}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle} \models B'(a)$, and by (ctype), $\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$. The converse can be shown if we start with the assumption $\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$. It is easy to extend the above reasoning to the case

$\mathcal{A}_1 = \{B(a)\}$ for a basic concept B over Σ_1, or $\mathcal{A}_1 = \{R(a,b)\}$ for a basic role R over Σ_1. As we quantify over all possible source ABoxes, we are free to choose any such concept B or role R, therefore, if $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, then for each basic concept or role X over Σ_1 and each basic concept or role X' over Σ_2,

$$\mathcal{T}_1 \cup \mathcal{T}_{12} \models X \sqsubseteq X' \quad \text{if and only if} \quad \mathcal{T}_2 \cup \mathcal{T}_{12} \models X \sqsubseteq X'.$$

This is the main intuition behind condition (ii) in Lemma 7.1.1.

In what follows, for better readability, for a source ABox $\mathcal{A}_1$, let us denote by $\mathcal{K}_1$ the KB $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, by $\mathcal{K}_2$ the KB $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$, and by $\mathcal{U}_1$ and $\mathcal{U}_2$ their respective canonical models. Then, for a TBox $\mathcal{T}$, a concept B and a role R, we write $\mathcal{T} \models B \mathrel{\underset{\rightsquigarrow}{\sqsubseteq}} \exists R$ if for an arbitrary constant $a \in N_a$, it holds that $a \rightsquigarrow_{\langle \mathcal{T}, \{B(a)\}\rangle} w_{[R]}$.

2) Again, let $\mathcal{A}_1 = \{A(a)\}$ for an atomic concept A over Σ_1, and assume that

$$\begin{aligned} \mathcal{T}_1 &\models A \mathrel{\underset{\rightsquigarrow}{\sqsubseteq}} \exists R && \text{for a basic role } R \text{ over } \Sigma_1,\\ \mathcal{T}_1 \cup \mathcal{T}_{12} &\models R \sqsubseteq R' && \text{for a basic role } R' \text{ over } \Sigma_2, \text{ and}\\ \mathcal{T}_1 \cup \mathcal{T}_{12} &\models \exists R^- \sqsubseteq B' && \text{for some basic concept } B' \text{ over } \Sigma_2. \end{aligned}$$

Then

$$aw_{[R]} \in \Delta^{\mathcal{U}_1}, \quad R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{U}_1}(a, aw_{[R]}), \quad \text{and} \quad B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_1}(aw_{[R]}).$$

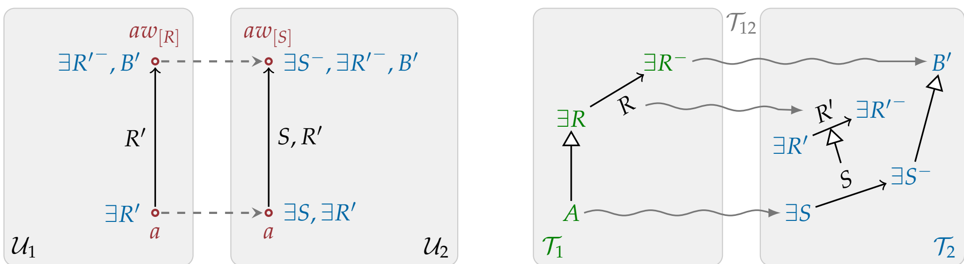
Next, for $\mathcal{T}_2$ to be a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, by Lemma 6.1.4, it follows that $\mathcal{U}_1$ has to be finitely Σ_2-homomorphically equivalent to $\mathcal{U}_2$. Let h be a Σ_2-homomorphism from all paths of length two in $\Delta^{\mathcal{U}_1}$ to $\Delta^{\mathcal{U}_2}$. Then $h(a) = a$ and there exists $aw_{[S]} \in \Delta^{\mathcal{U}_2}$ for a basic role S over Σ_2 (as $\mathsf{Ind}(\mathcal{A}_1) = \{a\}$ and there are no loops on a in $\mathcal{A}_1$, the image of $aw_{[R]}$ cannot be a constant) such that

$$h(aw_{[R]}) = aw_{[S]}, \quad R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{U}_2}(a, aw_{[S]}), \quad \text{and} \quad B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_2}(aw_{[S]}).$$

By construction of the canonical model and by the shape of $\mathcal{T}_{12}$, it follows then that it should be the case that

$$\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \sqsubseteq \exists S, \quad \mathcal{T}_2 \models S \sqsubseteq R', \quad \text{and} \quad \mathcal{T}_2 \models \exists S^- \sqsubseteq B'.$$

Then the Σ_2-reducts of the canonical models of $\mathcal{K}_1$ and $\mathcal{K}_2$, and inclusions implied by $\mathcal{T}_1$, $\mathcal{T}_{12}$, and $\mathcal{T}_2$ can be partially depicted as follows.



Clearly, given $\mathcal{T}_2$ and $\mathcal{T}_{12}$, one can check existence of such S effectively. Note that in the case, $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_1}(a, aw_{[R]}) = \emptyset$, the homomorphic image of $aw_{[R]}$ could be any element y in $\Delta^{\mathcal{U}_2}$ with $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_1}(aw_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_2}(y)$. This is the intuition behind condition (iii) in Lemma 7.1.1.

3) Continue with $\mathcal{A}_1 = \{A(a)\}$ for an atomic concept A over Σ_1, and assume now

$$\begin{aligned} \mathcal{T}_2 \cup \mathcal{T}_{12} \models A \mathrel{\underset{\sim}{\sqsubseteq}} \exists R' &\quad \text{for a basic role } R' \text{ over } \Sigma_2, \\ \mathcal{T}_2 \models \exists R'^- \sqsubseteq B' &\quad \text{for some basic concept } B' \text{ over } \Sigma_2. \end{aligned}$$

Then

$$aw_{[R']} \in \Delta^{\mathcal{U}_2}, \quad B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_2}(aw_{[R']}), \quad \text{and} \quad R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{U}_2}(a, aw_{[R']}).$$

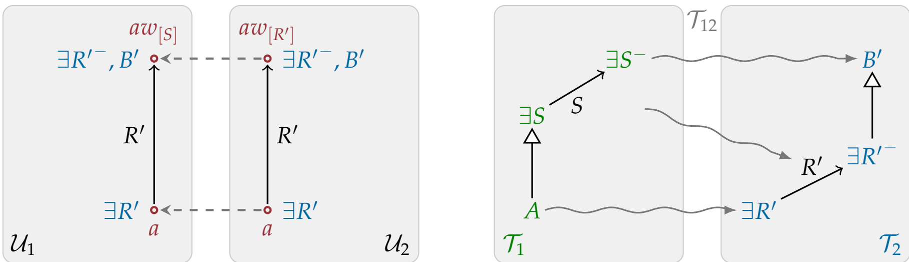
Now we are interested in homomorphisms in the opposite direction, so let h be a Σ_2-homomorphism from all paths of length two in $\Delta^{\mathcal{U}_2}$ to $\Delta^{\mathcal{U}_1}$. Then $h(a) = a$ and there exists $aw_{[S]} \in \Delta^{\mathcal{U}_1}$ for a basic role S such that

$$h(aw_{[R']}) = aw_{[S]}, \quad B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_1}(aw_{[S]}), \quad \text{and} \quad R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{U}_1}(a, aw_{[S]}).$$

By construction of the canonical model and by the shape of $\mathcal{T}_{12}$, it follows then that it should be the case that

$$\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq \exists S, \quad \mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R', \quad \text{and} \quad \mathcal{T}_1 \cup \mathcal{T}_2 \models \exists S^- \sqsubseteq B'.$$

Then the Σ_2-reducts of the canonical models of $\mathcal{K}_1$ and $\mathcal{K}_2$, and inclusions implied by $\mathcal{T}_1$, $\mathcal{T}_{12}$, and $\mathcal{T}_2$ can be partially depicted as follows.



Again, given $\mathcal{T}_1$ and $\mathcal{T}_{12}$, the existence of such S can be checked effectively. This is the intuition behind condition (iv) in Lemma 7.1.1.

Observe that it is sufficient to consider only chains of roles of length 1 as it was done in **2)** and **3)**. That is, if $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq \exists R$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists R^- \sqsubseteq \exists Q$ for roles R,Q, and for $\mathcal{A}_1 = \{A(a)\}$, $aw_{[R]}w_{[Q]} \in \Delta^{\mathcal{U}_1}$, it is enough to consider two separate cases covered by condition (iii) and discussed in **2)**:

- $\mathcal{A}_1 = \{A(a)\}$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq \exists R$,
- $\mathcal{A}_1 = \{\exists R^-(a)\}$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists R^- \sqsubseteq \exists Q$.

And analogously, for chains of roles by $\mathcal{T}_2 \cup \mathcal{T}_{12}$.

4) To conclude, we analyze the cases when $\mathcal{T}_1$, $\mathcal{T}_{12}$ and $\mathcal{T}_2$ contain disjointness assertions. First, notice that without loss of generality we can assume that there are no disjointness assertions in $\mathcal{T}_1$ as Equation (†) should be satisfied only for ABoxes $\mathcal{A}_1$ that are consistent with $\mathcal{T}_1$. So we will take into account only disjointness in $\mathcal{T}_{12}$ and $\mathcal{T}_2$. Next, for a source ABox $\mathcal{A}_1$ consistent with $\mathcal{T}_1$ it is possible that $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is inconsistent due to the disjointness assertions in the mapping, which will make all possible tuples to be in the answer to every query. Below we say a pair (B, B') of basic concepts is $\mathcal{T}$*-consistent* for a TBox $\mathcal{T}$, if the KB $\langle \mathcal{T}, \{B(a), B'(a)\} \rangle$ is consistent, where a is an arbitrary constant, and (B, B') is $\mathcal{T}$*-inconsistent* otherwise.

Consider a two-element ABox $\mathcal{A}_1 = \{A(a), C(a)\}$ for atomic concepts A, C over Σ_1, and assume it is consistent with $\mathcal{T}_1$. Further, assume (A, C) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, then $\mathcal{K}_1$ is inconsistent, and by definition of certain answers over an inconsistent KB, $cert(q, \mathcal{K}_1) = AllTup(q, \mathcal{A}_1)$ for each target UCQ q. Therefore, in order for $\mathcal{T}_2$ to be a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, $\mathcal{K}_2$ has to be inconsistent as well. To ensure that it is the case, we need to check that (A, C) is also $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent. Similarly in the opposite direction, if we start with the assumption (A, C) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, it should be verified that (A, C) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. This is the intuition behind condition (i). Observe that this condition guarantees that for every ABox $\mathcal{A}_1$ over Σ_1 that is consistent with $\mathcal{T}_1$, it holds that: $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent if and only $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent.

We introduce a couple of notations and prove several simple properties about *DL-Lite*$_\mathcal{R}$ KBs and their canonical models before proceeding with the characterization of UCQ-representations. Here, for a TBox $\mathcal{T}$, we say a pair (R, R') of basic roles is $\mathcal{T}$-*consistent*, if the KB $\langle \mathcal{T}, \{R(a,b), R'(a,b)\} \rangle$ is consistent, where a, b are arbitrary constants, and (R, R') is $\mathcal{T}$-*inconsistent* otherwise. Moreover, a concept or role X is $\mathcal{T}$-consistent iff (X, X) is $\mathcal{T}$-*consistent*, and $\mathcal{T}$-*inconsistent* otherwise.

Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a consistent KB, $a, b \in N_a$, $\sigma \in \Delta^{\mathcal{U}_\mathcal{K}}$, and $\mathsf{tail}(\sigma) \leadsto_\mathcal{K} w_{[R]}$. Then the following properties hold:

(CTYPE) $B \in \mathbf{t}^{\mathcal{U}_\mathcal{K}}(a)$ iff $\mathcal{A} \models B'(a)$ and $\mathcal{T} \models B' \sqsubseteq B$, and
$R \in \mathbf{r}^{\mathcal{U}_\mathcal{K}}(a,b)$ iff $\mathcal{A} \models R'(a,b)$ and $\mathcal{T} \models R' \sqsubseteq R$;

Proof: first, by definition of the canonical model, $B \in \mathbf{t}^{\mathcal{U}_\mathcal{K}}(a)$ if and only if $\mathcal{K} \models B(a)$. Next, assume $\mathcal{A} \not\models B(a)$, i.e., neither $B(a) \in \mathcal{A}$, nor $R(a,b) \notin \mathcal{A}$ for $B = \exists R$ and some $b \in N_a$. Obviously, $a \in \mathsf{Ind}(\mathcal{A})$, so for some concept B', $B'(a) \in \mathcal{A}$, or for some role R, $R(a,b) \in \mathcal{A}$. By contradiction, assume that $\mathcal{T} \not\models B' \sqsubseteq B$ for each $B'(a) \in \mathcal{A}$, and $\mathcal{T} \not\models \exists R \sqsubseteq B$ for each $R(a,b) \in \mathcal{A}$. Then there exists a model $\mathcal{I}$ of $\mathcal{K}$ such that $a^\mathcal{I} \notin B^\mathcal{I}$, which contradicts $\mathcal{K} \models B(a)$. Hence, $\mathcal{T} \models B' \sqsubseteq B$ for some $B'(a) \in \mathcal{A}$ or $\mathcal{T} \models \exists R \sqsubseteq B$ for some $R(a,b) \in \mathcal{A}$. The opposite direction is obvious. The proof for $R \in \mathbf{r}^{\mathcal{U}_\mathcal{K}}(a,b)$ is analogous.

(NTYPE) $B \in \mathbf{t}^{\mathcal{U}_\mathcal{K}}(\sigma w_{[R]})$ iff $\mathcal{T} \models \exists R^- \sqsubseteq B$, and
$R \in \mathbf{r}^{\mathcal{U}_\mathcal{K}}(\sigma, \sigma w_{[R']})$ iff $\mathcal{T} \models R' \sqsubseteq R$.

Proof: Follows from the definition of the canonical model and the types.

(CGEN) Let $a \leadsto_\mathcal{K} w_{[R]}$ for some basic role R. Then there exists a basic concept B, such that $\mathcal{A} \models B(a)$ and $\mathcal{T} \models B \sqsubseteq_{\leadsto} \exists R$.

Proof: by definition of $a \leadsto_\mathcal{K} w_{[R]}$ it follows that $\mathcal{K} \models \exists R(a)$ and R is a minimal with respect to $\leq_\mathcal{T}$ role among all $\{R' \mid \mathcal{K} \models \exists R'(a)\}$. By (ctype) we have that $\mathcal{A} \models B(a)$ for some concept B, and $\mathcal{T} \models B \sqsubseteq \exists R$. Now, consider KB $\mathcal{B} = \langle \mathcal{T}, \{B(o)\} \rangle$ for some $o \in N_a$. Obviously, $\mathcal{B} \models \exists R(o)$, $\mathcal{B} \not\models R(o,o)$, and R is a minimal with respect to $\leq_\mathcal{T}$ role among all $\{R' \mid \mathcal{B} \models \exists R'(o)\}$. Therefore, $o \leadsto_\mathcal{B} w_{[R]}$, and $\mathcal{T} \models B \sqsubseteq_{\leadsto} \exists R$.

(NGEN) Let $w_S \leadsto_\mathcal{K} w_{[R]}$ for basic roles S and R. Then $\mathcal{T} \models \exists S^- \sqsubseteq_{\leadsto} \exists R$.

Proof: by definition of $w_{[S]} \leadsto_\mathcal{K} w_{[R]}$ it follows that $\mathcal{T} \models \exists S^- \sqsubseteq \exists R$, $[S^-] \neq [R]$, and R is a minimal with respect to $\leq_\mathcal{T}$ role among all $\{R' \mid \mathcal{T} \models \exists S^- \sqsubseteq$

$\exists R'\}$. Consider KB $\mathcal{B} = \langle \mathcal{T}, \{\exists S^-(o)\}\rangle$ for some $o \in N_a$. The rest of the proof is similar to the proof of (cgen).

(STYPE) Let $\{B_1, \ldots, B_n\}$ be a set of basic concepts, and $\mathcal{T}'$ a TBox such that $\mathcal{B} = \langle \mathcal{T}, \{B_1(o), \ldots, B_n(o)\}\rangle$ and $\langle \mathcal{T} \cup \mathcal{T}', \mathcal{A}\rangle$ are consistent. Assume $y \in \Delta^{\mathcal{G}_\mathcal{B}}$. If for some $\sigma \in \Delta^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}$, $\{B_1, \ldots, B_n\} \subseteq \mathbf{t}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\sigma)$, then there exists $\delta \in \Delta^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}$ such that

$$\mathbf{t}^{\mathcal{G}_\mathcal{B}}(y) \subseteq \mathbf{t}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\delta) \quad \text{and} \quad \mathbf{r}^{\mathcal{G}_\mathcal{B}}(o, y) \subseteq \mathbf{r}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\sigma, \delta) \tag{13}$$

Proof: consider the cases of $y \in \Delta^{\mathcal{G}_\mathcal{B}}$. If $y = o$, then $\delta = \sigma$. Let $y = ow_{[R_1]} \cdots w_{[R_m]}$ for $m \geq 1$: then for some $1 \leq i \leq n$, $\mathcal{T} \models B_i \sqsubseteq \exists R_1$, and for $1 \leq j < m$, $\mathcal{T} \models \exists R_j^- \sqsubseteq \exists R_{j+1}$. Obviously, these entailments are valid in $\mathcal{T} \cup \mathcal{T}'$, so $\exists R_1 \in \mathbf{t}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\sigma)$ and there exists $\delta_1 \in \Delta^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}$ s.t. $R_1 \in \mathbf{r}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\sigma, \delta_1)$ and $\exists R_1^- \in \mathbf{t}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\delta_1)$. Moreover, for each $1 \leq j < m$, we have that $\exists R_{j+1} \in \mathbf{t}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\delta_j)$ and there exists $\delta_{j+1} \in \Delta^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}$ such that $R_{j+1} \in \mathbf{r}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\delta_j, \delta_{j+1})$ and $\exists R_{j+1}^- \in \mathbf{t}^{\mathcal{U}_{\langle \mathcal{T}\cup\mathcal{T}',\mathcal{A}\rangle}}(\delta_{j+1})$. So we take δ to be equal to δ_m. It is easy to see that (13) is satisfied.

Finally, we list all conditions that characterize UCQ-representations.

Lemma 7.1.1. *A TBox $\mathcal{T}_2$ over Σ_2 is a* UCQ*-representation of a TBox $\mathcal{T}_1$ over Σ_1 under the mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$* if and only if *following conditions hold:*

(i) *for each pair of $\mathcal{T}_1$-consistent concepts or roles X, X' over Σ_1, (X, X') is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent iff (X, X') is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent;*

(ii) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept or role X over Σ_1 and each X' over Σ_2, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models X \sqsubseteq X'$ iff $\mathcal{T}_2 \cup \mathcal{T}_{12} \models X \sqsubseteq X'$;*

(iii) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1 and each role R such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$ there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}$ such that*

(a) $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_2\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(y)$,

(b) $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_2\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y)$;

(iv) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1 and each role R such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$ there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}$ such that*

(a) $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_1\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(y)$,

(b) $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_1\cup\mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y)$

Proof. ($\Leftarrow$) Let the conditions above hold for $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_{12}$. Let $\mathcal{A}_1$ be an ABox over Σ_1 such that $\langle \mathcal{T}_1, \mathcal{A}_1\rangle$ is consistent, denote by $\mathcal{K}_1$ the KB $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$, and by $\mathcal{K}_2$ the KB $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1\rangle$. We show $\mathcal{K}_1$ is Σ_2-query equivalent to $\mathcal{K}_2$.

Observe that condition (i) ensures that $\mathcal{K}_1$ is consistent iff $\mathcal{K}_2$ is consistent. Indeed, if $\mathcal{K}_1$ is consistent, then for each pair of basic concepts B, B' over Σ_1 such that $\mathcal{A}_1 \models B(a)$ and $\mathcal{A}_1 \models B'(a)$ for some $a \in \mathsf{Ind}(\mathcal{A}_1)$, the KB ${\mathcal{K}_1}' = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \cup \{B(a), B'(a)\}\rangle$ is consistent, and by monotonicity of first-order logic we obtain that the KB $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(a), B'(a)\}\rangle$ is also consistent, and thus (B, B') is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. And similarly, for each pair of basic roles R, R' over Σ_1 such that $\mathcal{A}_1 \models R(b,c)$ and $\mathcal{A}_1 \models R'(b,c)$ for some $b, c \in \mathsf{Ind}(\mathcal{A}_1)$, we can derive that (R, R') is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Then, by (i) for each B, B' over Σ_1 such that $\mathcal{A}_1 \models B(a)$ and $\mathcal{A}_1 \models B'(a)$ for some $a \in \mathsf{Ind}(\mathcal{A}_1)$, (B, B') is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent, and for each R, R' over Σ_1 such that $\mathcal{A}_1 \models R(b,c)$ and $\mathcal{A}_1 \models R'(b,c)$ for some $b, c \in \mathsf{Ind}(\mathcal{A}_1)$, (R, R') is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent. To see that $\mathcal{K}_2$ is consistent, consider the interpretation $\mathcal{I}$ defined as the union of the canonical models $\mathcal{U}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(a), B'(a)\}\rangle}$ and $\mathcal{U}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{R(b,c), R'(b,c)\}\rangle}$ for B, B', R, R', and a, b, c as above, that is a model $\mathcal{K}_2$. Note that in this paragraph, B and B' can denote the same concept, and R and R' can denote the same role. The proof can be inverted to show $\mathcal{K}_2$ is consistent implies $\mathcal{K}_1$ is consistent.

First, assume $\mathcal{K}_1$ is inconsistent, it follows $\mathit{cert}(q, \mathcal{K}_1) = \mathit{AllTup}(q)$ for each UCQ q over Σ_2. By the paragraph above, $\mathcal{K}_2$ is inconsistent, so $\mathit{cert}(q, \mathcal{K}_2) = \mathit{AllTup}(q)$ for each UCQ q over Σ_2 as well, hence $\mathcal{K}_1$ is Σ_2-query equivalent to $\mathcal{K}_2$.

Now assume $\mathcal{K}_1$ is consistent. In Proposition 7.1.2 below, we show that from (ii) and (iii) it follows that $\mathcal{U}_{\mathcal{K}_1}$ is Σ_2-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_2}$. Since $\mathcal{K}_2$ is consistent, then we can apply Lemma 6.1.4 to obtain $\mathcal{K}_2$ Σ_2-query entails $\mathcal{K}_1$. On the other hand, in Proposition 7.1.3, we show that (ii) and (iv) imply that $\mathcal{U}_{\mathcal{K}_2}$ is Σ_2-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_1}$, and $\mathcal{K}_1$ Σ_2-query entails $\mathcal{K}_2$ by Lemma 6.1.4. We again obtain $\mathcal{K}_1$ is Σ_2-query equivalent to $\mathcal{K}_2$.

($\Rightarrow$) Assume, by contradiction, one of the conditions (i) – (iv) is not satisfied. We produce a $\mathcal{T}_1$-consistent ABox $\mathcal{A}_1$ over Σ and a Boolean CQ q over Σ_2 such that it is not the case that $\mathcal{K}_1 \models q$ iff $\mathcal{K}_2 \models q$.

Assume, first, condition (i) is violated, then we take $\mathcal{A}_1 = \{B_1(o), B_2(o)\}$ for concepts B_1 and B_2 violating it and $q = B_1(a)$ for some constant a distinct from o. If B_1, B_2 are $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, but $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, it follows $\mathcal{K}_1 \not\models q$ and $\mathcal{K}_2 \models q$, and the opposite holds if B_1, B_2 are $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent, but $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. If (i) is violated for roles, the proof is analogous.

Let now condition (ii) be violated for some $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1. Assume there is B' such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models B \sqsubseteq B'$, and consider $\mathcal{A}_1 = \{B(o)\}$ and $q = B'(o)$. Then by $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(o)$ and $B' \notin \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(o)$, so it follows $\mathcal{U}_{\mathcal{K}_1} \models q$ and $\mathcal{U}_{\mathcal{K}_2} \not\models q$; finally by Lemma 6.1.1 it follows $\mathcal{K}_1 \models q$ and $\mathcal{K}_2 \not\models q$. The opposite follows if we assume $\mathcal{T}_1 \cup \mathcal{T}_{12} \not\models B \sqsubseteq B'$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$, which completes the proof for this case. If (ii) is violated for some role, the proof is analogios.

Next, assume condition (iii) is violated, so there exists a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1 and a role R such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$ and for $\mathcal{A}_1 = \{B(o)\}$, for all $y \in \Delta^{\mathcal{G}_{\mathcal{K}_2}}$ either

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(w_{[R]}) \not\subseteq \mathbf{t}^{\mathcal{G}_{\mathcal{K}_2}}(y) \quad \text{or} \quad \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(o, w_{[R]}) \not\subseteq \mathbf{r}^{\mathcal{G}_{\mathcal{K}_2}}(o, y).$$

Let $\vec{B} = \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(w_{[R]})$, $\vec{R} = \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(o, w_{[R]})$, and consider

$$q = \exists x \Big(\bigwedge_{B' \in \vec{B}} B'(x) \wedge \bigwedge_{R' \in \vec{R}} R'(o, x) \Big),$$

where $B'(x)$ denotes atom $A(x)$ if $B' = A$ for atomic concept A, and $B'(x)$ denotes formula $\exists x'.S(x, x')$ if $B' = \exists S$ for a role S. Then $\mathcal{U}_{\mathcal{K}_1} \models q$ with substitution $x \mapsto ow_{[R]}$. On the other hand, $\mathcal{U}_{\mathcal{K}_2} \not\models q$ as there exists no substitution for x in $\Delta^{\mathcal{U}_{\mathcal{K}_2}}$. Using Lemma 6.1.1 we then obtain $\mathcal{K}_1 \models q$ and $\mathcal{K}_2 \not\models q$.

The case when (iv) is violated is analogous to the case above. The proof is complete. In the end we provide the proofs of the aforementioned propositions.

Proposition 7.1.2. *Let conditions (ii) and (iii) hold, and $\mathcal{A}_1$ an ABox over Σ_1 such that $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ and $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ are consistent. Then $\mathcal{U}_{\mathcal{K}_1}$ is Σ_2-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_2}$.*

Proof. We build a function h from $\Delta^{\mathcal{U}_{\mathcal{K}_1}}$ to $\Delta^{\mathcal{U}_{\mathcal{K}_2}}$, which is a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{K}_1}$ to $\mathcal{U}_{\mathcal{K}_2}$.

Base of induction. Initially, for each $a \in \mathsf{Ind}(\mathcal{A}_1)$ we define $h(a) = a$. Let us immediately verify that $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a)$. Let $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a)$, it follows by (ctype) there exists B over Σ_1 such that $\mathcal{A}_1 \models B(a)$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$. Note that B is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, then by (ii), $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$, therefore we obtain $B' \in \mathbf{t}^{\mathcal{U}_{\mathcal{K}_2}}(a)$. The proof of $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a,b) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a,b)$ is analogous.

Next, assume $\sigma \in \Delta^{\mathcal{U}_{\mathcal{K}_1}}$ and $\sigma = aw_{[R]}$, we show how to define $h(\sigma)$. It follows $a \rightsquigarrow_{\mathcal{K}_1} w_{[R]}$ and by (cgen) we obtain a concept B over Σ_1 such that $\mathcal{A}_1 \models B(a)$, and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$. Then B is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, and by (iii) there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o,y).$$

Since $\{B\} \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}_2}}(a)$, by (stype) there exists $\delta \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ such that

$$\mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}_2}}(\delta), \quad \text{and} \quad \mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o,y) \subseteq \mathbf{r}^{\mathcal{U}_{\mathcal{K}_2}}(a,\delta).$$

As for a TBox $\mathcal{T}$, ABoxes $\mathcal{A}$ and $\mathcal{A}'$, and $x \in \Delta^{\mathcal{G}_{\langle \mathcal{T}, \mathcal{A}\rangle}}$, $z \in \Delta^{\mathcal{U}_{\langle \mathcal{T}, \mathcal{A}'\rangle}}$ with $x = \mathsf{tail}(z)$, the concept and role types of x and z coincide, it follows now by transitivity that

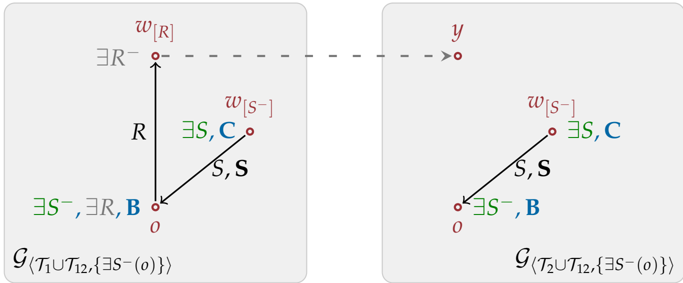
$$\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(aw_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(\delta), \quad \text{and} \quad \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a, aw_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a,\delta).$$

Hence, we assign $h(\sigma) = \delta$.

Inductive step. We show now how to define homomorphism for $\sigma w_{[R]} \in \Delta^{\mathcal{U}_{\mathcal{K}_1}}$ with $\sigma = \sigma' w_{[S]}$ given that the $h(\sigma)$ and $h(\sigma')$ are defined. It follows $w_{[S]} \rightsquigarrow_{\mathcal{K}_1} w_{[R]}$ and S is a basic role over Σ_1 by the structure of $\mathcal{T}_1 \cup \mathcal{T}_{12}$. Then $\exists S^-$ is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, and by (ngen), $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists S^- \sqsubseteq_{\rightsquigarrow} \exists R$. So (iii) is triggered, and there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}$ satisfying

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(y),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(o,y).$$

Let $\mathbf{B} = \mathbf{u}_{\Sigma_2}^{\mathcal{T}_{12}}(\exists S^-)$, $\mathbf{C} = \mathbf{u}_{\Sigma_2}^{\mathcal{T}_{12}}(\exists S)$, and $\mathbf{S} = \mathbf{u}_{\Sigma_2}^{\mathcal{T}_{12}}(S)$, where for a TBox $\mathcal{T}$ and a concept B, $\mathbf{u}_{\Sigma}^{\mathcal{T}}(B)$ denotes the set of all concepts B' over Σ such that $\mathcal{T} \models B \sqsubseteq B'$; and $\mathbf{u}_{\Sigma}^{\mathcal{T}}(R)$ is defined analogously for a role R. Then $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}$ and $\mathcal{U}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}$ can be partially depicted as follows. Note that here the presented concept and role labels are NOT the exact concept and role types. Moreover, we depict only those individuals and links between them, which are guaranteed to exist given the information at hand. Note also, that in the pictures further in this proof, we depict only the necessary bits of information.



Denote by $\mathbf{B}(o)$ assertions $B_1(o), \dots, B_m(o)$ for $B_i \in \mathbf{B}$, and similarly for $\mathbf{C}(a)$. Moreover, denote by $\mathbf{S}(a,o)$ assertions $S_1(a,o), \dots, S_k(a,o)$ for $S_i \in \mathbf{S}$. There are two possible cases considering that $\mathcal{T}_{12}$ is a set of inclusions from Σ_1 to Σ_2, $\mathcal{T}_2$ is a TBox over Σ_2, and S is a role over Σ_1.

(I) $o \rightsquigarrow_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle} w_{[Q_1]} \rightsquigarrow \cdots \rightsquigarrow w_{[Q_n]}$, $n \geq 0$ and Q_i are roles over Σ_2. Then, if $n = 0$, $y = o$, otherwise $y = w_{[Q_n]}$.

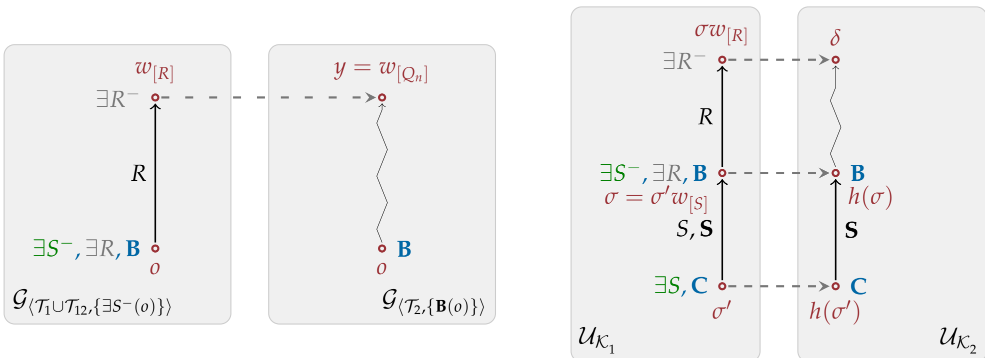
Consider KB $\langle \mathcal{T}_2, \{\mathbf{B}(o)\}\rangle$, then we obtain that $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{B}(o)\}\rangle}}$ and

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(y) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{B}(o)\}\rangle}}(y),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(o,y) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{B}(o)\}\rangle}}(o,y).$$

Observe that $\mathbf{B} \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma))$, since obviously $\mathbf{B} \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(\sigma)$ and h is a homomorphism on σ. Therefore, by (stype) we obtain $\delta \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{B}(o)\}\rangle}}(y) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(\delta), \quad \text{and} \quad \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{B}(o)\}\rangle}}(o,y) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma),\delta).$$

As above, it follows $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(\sigma w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(\delta)$, and $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(\sigma, \sigma w_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma),\delta)$. Hence, we assign $h(\sigma w_{[R]}) = \delta$. This case can be depicted as follows:



(II) $o \rightsquigarrow_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle} w_{[S^-]} \rightsquigarrow w_{[Q_1]} \rightsquigarrow \cdots \rightsquigarrow w_{[Q_n]}$, $n \geq 0$, Q_i are roles over Σ_2. Then, if $n = 0$, $y = w_{[S^-]}$, otherwise $y = w_{[Q_n]}$.

Consider KB $\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle$. Then $a \rightsquigarrow_{\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle} w_{[Q_1]} \rightsquigarrow \cdots \rightsquigarrow w_{[Q_n]}$, $y' \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle}}$: if $n = 0$, $y' = a$, otherwise $y' = w_{[Q_n]}$, and

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(y) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle}}(y'),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(o,y) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle}}(o,y').$$

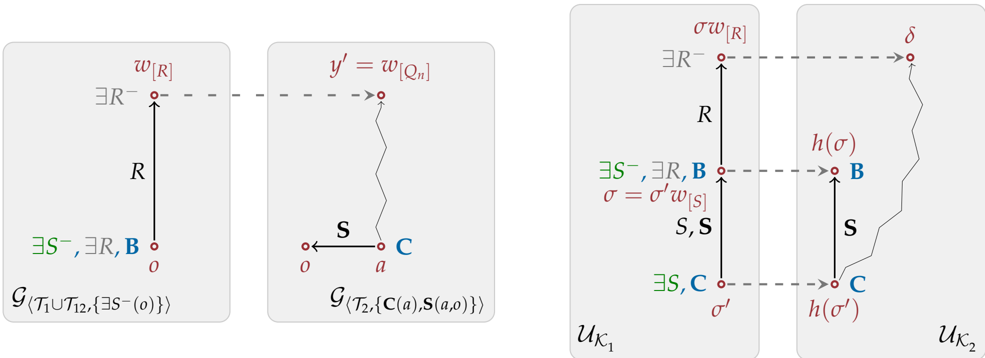
As above, $\mathbf{C} \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma'))$, therefore by (stype) we obtain $\delta \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ such that $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle}}(y') \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(\delta)$.

Observe that if $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(o, w_{[R]}) \neq \emptyset$, it has to be the case that

$$y = w_{[S^-]}, \quad y' = a, \quad \text{and} \quad \delta = h(\sigma').$$

Let $R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}(o, w_{[R]})$, it follows $R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2, \{\mathbf{C}(a), \mathbf{S}(a,o)\}\rangle}}(o, a)$, and from the latter, $\mathcal{T}_2 \models S_i^- \sqsubseteq R'$ for some $S_i \in \mathbf{S}$. As $S_i \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma'), h(\sigma))$, we obtain that $R' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma), h(\sigma'))$.

All in all, it follows that $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(\sigma w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(\delta)$, and $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(\sigma, \sigma w_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(h(\sigma), \delta)$. Hence, we set $h(\sigma w_{[R]}) = \delta$. We conclude with a graphical representation of this case:



In such a way we can define $h(\sigma)$ for each $\sigma \in \Delta^{\mathcal{U}_{\mathcal{K}_1}}$, hence h is a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{K}_1}$ to $\mathcal{U}_{\mathcal{K}_2}$. □

Proposition 7.1.3. *Let conditions (ii) and (iv) hold, and $\mathcal{A}_1$ an ABox over Σ_1 such that $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ and $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ are consistent. Then $\mathcal{U}_{\mathcal{K}_2}$ is Σ_2-homomorphically embeddable into $\mathcal{U}_{\mathcal{K}_1}$.*

Proof. We build a function h from $\Delta^{\mathcal{U}_{\mathcal{K}_2}}$ to $\Delta^{\mathcal{U}_{\mathcal{K}_1}}$, which is a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{K}_2}$ to $\mathcal{U}_{\mathcal{K}_1}$.

Base of induction. Initially, for each $a \in \mathsf{Ind}(\mathcal{A}_1)$ we define $h(a) = a$. Let us immediately verify that $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a)$. Let $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a)$, it follows by (ctype) there exists B over Σ_1 such that $\mathcal{A}_1 \models B(a)$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$. Then B is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent (recall that $\mathcal{K}_1 = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent), so by (ii), $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$, therefore we obtain $B' \in \mathbf{t}^{\mathcal{U}_{\mathcal{K}_1}}(a)$. The proof of $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a, b) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a, b)$ is analogous.

Next, assume $\sigma \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ and $\sigma = a w_{[R]}$, we show how to define $h(\sigma)$. It follows $a \rightsquigarrow_{\mathcal{K}_2} w_{[R]}$ and by (cgen) we obtain B over Σ_1 such that $\mathcal{A}_1 \models B(a)$, and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$. We are going to show now there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y), \text{ and} \tag{14}$$

$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y). \tag{15}$$

Assume, first, R is a role over Σ_2, and observe that B is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, then by (iv) there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ satisfying (14) and (15).

Assume now R is a role over Σ_1, then it follows $B = \exists R$. Let $o \rightsquigarrow_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists R(o)\}\rangle} w_{[Q]}$ for a role Q over Σ_1 such that $\mathcal{T}_1 \models Q \sqsubseteq R$ (such Q always exists, for instance R itself if it does not have proper subroles). Then we choose y to be $w_{[Q]}$, and show first that (14) is satisfied. Let $B \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists R(o)\}\rangle}}(w_{[R]})$, then by (ntype), $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \exists R^- \sqsubseteq B$, and as $\exists R^- \in \mathbf{t}_{\Sigma_1}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists R(o)\}\rangle}}(w_{[Q]})$, by (ii) we obtain that $B \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists R(o)\}\rangle}}(w_{[Q]})$. In a similar way, we can show that (15) is satisfied.

To continue the proof consider $\{B\} \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}_1}}(a)$, then by (stype) there exists $\delta \in \Delta^{\mathcal{U}_{\mathcal{K}_1}}$ such that $\mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}_1}}(\delta)$ and $\mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y) \subseteq \mathbf{r}^{\mathcal{U}_{\mathcal{K}_1}}(a, \delta)$. It follows now using (14) that $\mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(aw_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(\delta)$. Analogously using (15) one obtains $\mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_2}}(a, aw_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(a, \delta)$.

Inductive step. We show how to define homomorphism for $\sigma w_{[R]} \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$ with $\sigma = \sigma' w_{[S]}$ given that $h(\sigma)$ is defined. It follows $w_{[S]} \rightsquigarrow_{\mathcal{K}_2} w_{[R]}$, therefore $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \exists S^- \sqsubseteq \exists R$, and R is a role over Σ_2 *distinct from* S^-. By (ntype) it also follows $\exists R \in \mathbf{t}^{\mathcal{U}_{\mathcal{K}_2}}(\sigma)$, and since h is a Σ_2-homomorphism, $\exists R \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\mathcal{K}_1}}(h(\sigma))$. As $\mathcal{A}_1$ is an ABox over Σ_1 and $\mathcal{T}_1$ is a TBox over Σ_1, there exists a concept B over Σ_1 such that $B \in \mathbf{t}^{\mathcal{U}_{\mathcal{K}_1}}(h(\sigma))$ and $\mathcal{T}_{12} \models B \sqsubseteq \exists R$. Next, assume that $o \rightsquigarrow_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle} w_{[Q]}$ for some role Q such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models Q \sqsubseteq R$. Then B is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists Q$. As above for $\sigma = aw_{[R]}$, by (iv) there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(w_{[Q]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y), \text{ and}$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, w_{[Q]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y).$$

Again, by (stype) we obtain δ in $\Delta^{\mathcal{U}_{\mathcal{K}_1}}$ such that $\mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}_1}}(\delta)$ and $\mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y) \subseteq \mathbf{r}^{\mathcal{U}_{\mathcal{K}_1}}(h(\sigma), \delta)$. Observe that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models Q \sqsubseteq R$, so the concept and role types of $w_{[R]}$ and $(o, w_{[R]})$ are subsumed by those of $w_{[Q]}$ and $(o, w_{[Q]})$ in $\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}$. Finally, we obtain that $\mathbf{t}^{\mathcal{U}_{\mathcal{K}_2}}(\sigma w_{[R]}) \subseteq \mathbf{t}^{\mathcal{U}_{\mathcal{K}_1}}(\delta)$ and $\mathbf{r}^{\mathcal{U}_{\mathcal{K}_2}}(\sigma, \sigma w_{[R]}) \subseteq \mathbf{r}^{\mathcal{U}_{\mathcal{K}_1}}(h(\sigma), \delta)$. Hence, we assign $h(\sigma w_{[R]}) = \delta$.

In such a way we can define $h(\sigma)$ for each $\sigma \in \Delta^{\mathcal{U}_{\mathcal{K}_2}}$, hence h is a Σ_2-homomorphism from $\mathcal{U}_{\mathcal{K}_2}$ to $\mathcal{U}_{\mathcal{K}_1}$. $\square$

This concludes the proof of Lemma 7.1.1. $\square$

Having devised a characterization of UCQ-representations, we discuss several examples of (non-)UCQ-representations.

Example 7.1.4. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot, \cdot)\}$, $\Sigma_2 = \{A'(\cdot), R'(\cdot, \cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{A \sqsubseteq A', \exists R^- \sqsubseteq B'\}$. Moreover, let $\mathcal{T}_1 = \{A \sqsubseteq \exists R\}$, and

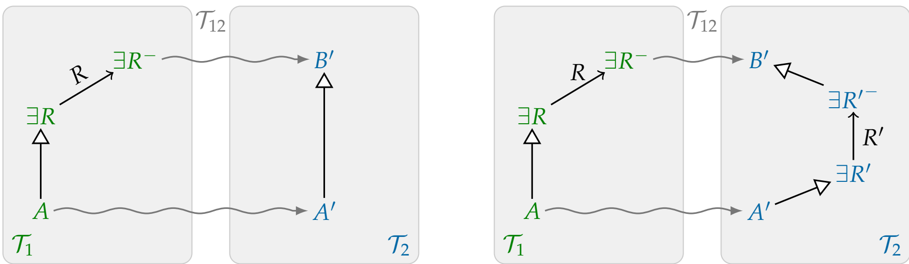
(1) $\mathcal{T}_2 = \{A' \sqsubseteq B'\}$. In Example 4.3.2 we showed that $\mathcal{T}_2$ is not a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. In fact, in this case condition (ii) is not satisfied, as $\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$ while $\mathcal{T}_1 \cup \mathcal{T}_{12} \not\models A \sqsubseteq B'$.

(2) $\mathcal{T}_2 = \{A' \sqsubseteq \exists R', \exists R'^- \sqsubseteq B'\}$. We also showed in that example that this $\mathcal{T}_2$ is not a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. In this case, condition (iv) is not satisfied, as $\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \underset{\rightsquigarrow}{\sqsubseteq} \exists R'$, but there exists no $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(w_{[R']}) \subseteq \mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(y),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(o, w_{[R']}) \subseteq \mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(o, y),$$

since neither $y = o$, nor $y = w_{[R]}$ in $\Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$ satisfy $R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(o, y)$.

These cases can be depicted in the following ER diagrams:



(1) (2)

■

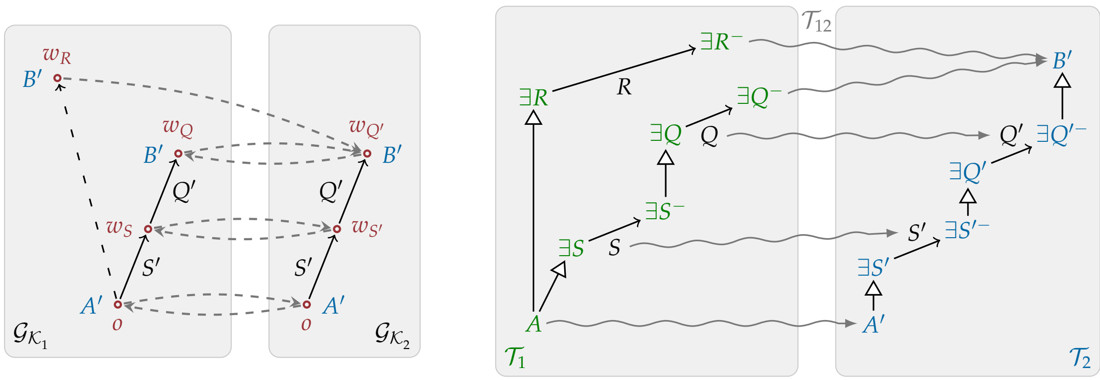
Example 7.1.5. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where

$$\begin{aligned} \Sigma_1 &= \{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot), Q(\cdot,\cdot)\} \\ \Sigma_2 &= \{A'(\cdot), B'(\cdot), S'(\cdot,\cdot), Q'(\cdot,\cdot)\} \\ \mathcal{T}_{12} &= \{A \sqsubseteq A', \exists R^- \sqsubseteq B', \\ &\qquad S \sqsubseteq S', Q \sqsubseteq Q', \exists Q^- \sqsubseteq B'\} \end{aligned} \quad \text{and let} \quad \begin{aligned} \mathcal{T}_1 &= \{A \sqsubseteq \exists R, \\ &\qquad A \sqsubseteq \exists S, \exists S^- \sqsubseteq \exists Q\} \\ \mathcal{T}_2 &= \{A' \sqsubseteq \exists S', \exists S'^- \sqsubseteq \exists Q', \\ &\qquad \exists Q'^- \sqsubseteq B'\} \end{aligned}$$

Then $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. We verify that conditions (iii) and (iv) are satisfied. First, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \underset{\rightsquigarrow}{\sqsubseteq} \exists R$: we take $w_{Q'} \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$ and it is easy to see that the following is satisfied:

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(w_R) \subseteq \mathbf{t}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(w_{Q'}),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(o, w_R) \subseteq \mathbf{r}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(o, w_{Q'}),$$

as $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(o, w_R) = \emptyset$. Then for each of $A \underset{\rightsquigarrow}{\sqsubseteq} \exists S$ and $A \underset{\rightsquigarrow}{\sqsubseteq} \exists Q$ entailed by $\mathcal{T}_1 \cup \mathcal{T}_{12}$, it should be clear that we take $w_{S'}$ and $w_{Q'}$ in $\Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$ and $\Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}$ respectively to satisfy condition (iii). As for the opposite direction, now differently from Example 7.1.4, for both $w_{S'}$ and $w_{Q'}$ in $\Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$ and $\Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}$ respectively, there exist w_S and w_Q in $\Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$ and $\Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{\exists S^-(o)\}\rangle}}$ that satisfy condition (iv). Below we illustrate the Σ_2-reducts of $\mathcal{G}_{\mathcal{K}_1}$ and $\mathcal{G}_{\mathcal{K}_2}$ for $\mathcal{K}_1 = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle$ and $\mathcal{K}_2 = \langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\}\rangle$ (concept labels of the form $\exists P, \exists P^-$ for a role P are not shown), and the ER diagrams of $\mathcal{T}_1$, $\mathcal{T}_{12}$ and $\mathcal{T}_2$:



Observe that in this example, if we remove axioms $A \sqsubseteq \exists R$ and $\exists R^- \sqsubseteq B'$ from $\mathcal{T}_1$ and $\mathcal{T}_{12}$, $\mathcal{T}_2$ will still be a UCQ-representation for the remaining $\mathcal{T}_1$ under the remaining $\mathcal{M}$. It means that an "image" of $w_R \in \Delta^{\mathcal{G}_{\mathcal{K}_1}}$ should be some $y \in \Delta^{\mathcal{G}_{\mathcal{K}_2}}$ for $\mathcal{K}_1 = \langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\} \rangle$ and $\mathcal{K}_2 = \langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{A(o)\} \rangle$, such that for $n \geq 1$

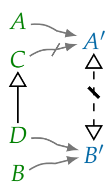
$$o \rightsquigarrow_{\mathcal{K}_2} y_1 \rightsquigarrow_{\mathcal{K}_2} \cdots \rightsquigarrow_{\mathcal{K}_2} y_n = y,$$

which is also a "one-to-one image" of some $x \in \Delta^{\mathcal{G}_{\mathcal{K}_1}}$, that is

$$o \rightsquigarrow_{\mathcal{K}_1} x_1 \rightsquigarrow_{\mathcal{K}_1} \cdots \rightsquigarrow_{\mathcal{K}_1} x_n = x,$$

$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(o, x_1) = \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_2}}(o, y_1)$, $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(x_i, x_{i+1}) = \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_2}}(y_i, y_{i+1})$, and $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_1}}(x_i) = \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\mathcal{K}_2}}(y_i)$. This fact is important for solving the non-emptiness problem in the next section. ∎

Example 7.1.6. Assume that $\mathcal{M} = (\{A(\cdot), B(\cdot), C(\cdot), D(\cdot)\}, \{A'(\cdot), B'(\cdot)\}, \mathcal{T}_{12})$, where $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq \neg A', D \sqsubseteq B'\}$, moreover, let $\mathcal{T}_1 = \{D \sqsubseteq C\}$ and $\mathcal{T}_2 = \{A' \sqsubseteq \neg B'\}$. As we showed in Example 4.3.4, $\mathcal{T}_2$ is not a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$. We verify that using the characterization. In fact, although, $\mathcal{T}_2$ satisfies condition (i) for the pair of concepts (A, D), which is both $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent and $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, $\mathcal{T}_2$ violates this condition for the pair (A, B), which is clearly $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, however $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent as $\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \sqsubseteq \neg B'$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$.



∎

Note that the proof of Lemma 7.1.1 implies another characterization of UCQ-representations, in the spirit of the characterizations of universal solutions and universal UCQ-solutions.

Lemma 7.1.7. *A TBox $\mathcal{T}_2$ over Σ_2 is a* UCQ-*representation of a TBox $\mathcal{T}_1$ over Σ_1 under a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$* if and only if *following conditions hold:*

- *for each ABox $\mathcal{A}_1$ consistent with $\mathcal{T}_1$, $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent iff $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ is consistent;*

- *for each ABox $\mathcal{A}_1$ consistent with $\mathcal{T}_1 \cup \mathcal{T}_{12}$, $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is Σ_2-homomorphically equivalent to $\mathcal{U}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$.*

Proof. ($\Rightarrow$) Follows from the proof of Lemma 7.1.1: if $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$ then conditions (i) – (iv) of Lemma 7.1.1 are satisfied, and the required follows from them.

($\Leftarrow$) Straightforward. □

Finally, we conclude with the main result of this section. We can derive an efficient algorithm for checking the membership problem for UCQ-representations from the conditions in Lemma 7.1.1. Combining it with the complexity of reasoning in *DL-Lite*$_\mathcal{R}$, we obtain the following complexity bound:

Theorem 7.1.8. *The membership problem for* UCQ*-representations is* NLOGSPACE-*complete.*

Proof. The lower bound can be obtained by the reduction from the directed graph reachability problem, which is known to be NLOGSPACE-hard: given a graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ and a pair of vertices $v_k, v_m \in \mathsf{V}$, decide if there is a directed path from v_k to v_m. To encode the problem, we need a source signature Σ_1 of concept names $\{V_i \mid v_i \in \mathsf{V}\}$ and a target signature Σ_2 of concept names $\{V_i' \mid v_i \in \mathsf{V}\}$. Consider $\mathcal{T}_1 = \{V_k \sqsubseteq V_m\} \cup \{V_i \sqsubseteq V_j \mid (v_i, v_j) \in \mathsf{E}\}$, $\mathcal{T}_{12} = \{V_i \sqsubseteq V_i' \mid v_i \in \mathsf{V}\}$, and $\mathcal{T}_2 = \{V_i' \sqsubseteq V_j' \mid (v_i, v_j) \in \mathsf{E}\}$. One can easily verify that the condition ii of Lemma 7.1.1 is satisfied iff there is a directed path from v_k to v_m in G, whereas the other conditions of Lemma 7.1.1 are satisfied trivially. Therefore,

- there is a directed path from v_k to v_m in G iff $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M} = (\Sigma, \Sigma_2, \mathcal{T}_{12})$.

This concludes the proof of the lower bound.

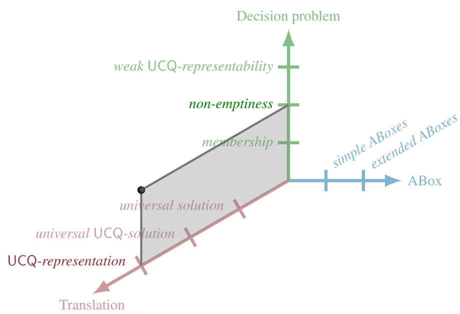
For the upper bound, we show that the conditions (i) – (iv) of Lemma (7.1.1) can be verified in NLOGSPACE. It is well known (see, e.g., [15]), that given a pair of *DL-Lite*$_\mathcal{R}$ concepts B, B', and a TBox $\mathcal{T}$, it can be verified in NLOGSPACE, if B, B' is $\mathcal{T}$ consistent (using an algorithm, based on directed graph reachability solving procedure); the same holds for a pair of *DL-Lite*$_\mathcal{R}$ roles R, R'. The same algorithm can be straightforwardly adopted to check, if $\mathcal{T} \models B \sqsubseteq B'$ or $\mathcal{T} \models R \sqsubseteq R'$. Therefore, clearly, the conditions (i) and (ii) can be verified in NLOGSPACE.

The conditions (iii) and (iv) are slightly more involved; first of all, observe that, given a concept B and a role R, and a TBox $\mathcal{T}$, it can be checked in NLOGSPACE, whether $\mathcal{T} \models B \sqsubseteq_{\rightsquigarrow} \exists R$, using an algorithm based on the directed graph reachability solving procedure. At the same time, given $z \in \{o\} \cup \{w_{[R]} \mid R \text{ is a role}\}$, we can verify, if there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}$ with $z = \mathsf{tail}(y)$: we "follow" the sequence of roles $R_1, \ldots, R_n = R$ (with $n \geq 0$) in the way that when we "guess" R_{i+1}, we check $w_{[R_i]} \rightsquigarrow_{\langle \mathcal{T}, \{B(o)\}\rangle} w_{[R_{i+1}]}$ (by the algorithm, similar to the one for checking $o \rightsquigarrow_{\langle \mathcal{T}, \{B(o)\}\rangle} w_{[R]}$), and "forget" R_i.

Furthermore, in a way similar to testing $\mathcal{T} \models B \sqsubseteq B'$, one can check if a concept $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{U}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(ow_{[R]})$ in NLOGSPACE; the same holds for checking if a role $R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(o, w_{[R]})$, and, then, for checking $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(y)$, for y as above. By combining the algorithms outlined above, one can produce a procedure that checks the conditions (iii) and (iv) in NLOGSPACE. □

7.2 THE NON-EMPTINESS PROBLEM

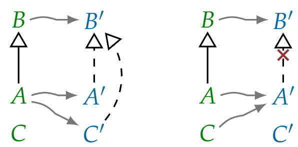
In this section we develop a bit more involved graph-theoretic techniques and show that the non-emptiness problem for UCQ-representations can also be solved efficiently (in NLogSpace).



We start with a example that provides some intuition to how the non-emptiness problem is solved.

Example 7.2.1. Assume $\mathcal{M}$ and a UCQ-representable $\mathcal{T}_1$ from Example 4.3.1-(3): $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot), C(\cdot)\}$, $\Sigma_2 = \{A'(\cdot), B'(\cdot), C'(\cdot)\}$ and $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', A \sqsubseteq C'\}$, and $\mathcal{T}_1 = \{A \sqsubseteq B\}$. It follows that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$. First and obvious requirement for a UCQ-representation $\mathcal{T}_2$ is that $\mathcal{T}_2$ should entail an axiom of the form $D' \sqsubseteq B'$ so that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$ (hence, $\mathcal{T}_{12} \models A \sqsubseteq D'$). On the other hand, it could be that $\mathcal{T}_{12} \models D \sqsubseteq D'$ for some D distinct from A, in which case it follows also $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D \sqsubseteq B'$. Since we want $\mathcal{T}_2$ to be a UCQ-representation, it should be the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq B'$. In this case, we can take D' equal to A' or C', and there exists no concept D: it is easy to see that there are two UCQ-representations of $\mathcal{T}_1$ under $\mathcal{M}$: $\{A' \sqsubseteq B'\}$ and $\{C' \sqsubseteq B'\}$.

Assume now a slightly different $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq A'\}$ from Example 4.3.1-(4), where we showed $\mathcal{T}_1$ is not UCQ-representable. As before, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq B'$. However now, the only candidate for D' is A', and there exists a concept C distinct from A such that $\mathcal{T}_{12} \models C \sqsubseteq A'$. So on the one hand, the only way to have a UCQ-representation $\mathcal{T}_2$ is to include axiom $A' \sqsubseteq B'$ to $\mathcal{T}_2$, but on the other hand since $\mathcal{T}_1 \cup \mathcal{T}_{12} \not\models C \sqsubseteq B'$, this axiom cannot be in $\mathcal{T}_2$. In general, there is no way to "represent" inclusion $A \sqsubseteq B'$ in the target, so in this case $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$.



(3) (4)

■

Example 7.2.2. Assume $\mathcal{M}$ and $\mathcal{T}_1$ from Example 4.3.4, where we showed $\mathcal{T}_1$ is not UCQ-representable: $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot), C(\cdot), D(\cdot)\}$, $\Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq \neg A', D \sqsubseteq B'\}$, and $\mathcal{T}_1 = \{D \sqsubseteq C\}$.

It follows that the pair of concepts (A, D) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent as $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq A'$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \neg A'$. So a candidate UCQ-representation $\mathcal{T}_2$ should be such that (A, D) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent. One possible way to achieve that is by having $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D \sqsubseteq \neg A'$, and since D is transferred only to B' through the mapping, it means that $\mathcal{T}_2$ should entail $B' \sqsubseteq \neg A'$, or $B' \sqsubseteq \neg B'$, or $A' \sqsubseteq \neg A'$. In the first case, however, we result with the pair (A, B) being $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent as well, since $A \sqsubseteq A'$ and $B \sqsubseteq B'$ are in $\mathcal{T}_{12}$. Then, for $\mathcal{T}_2$ to be a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, (A, B) should be $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which is not the case. So it cannot be that $\mathcal{T}_2 \models B' \sqsubseteq \neg A'$. In the second case, we result with the pair (B, B) being $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, and since (B, B) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent, it cannot be that $\mathcal{T}_2 \models B' \sqsubseteq \neg B'$. Similarly, we get that it cannot be the case that $\mathcal{T}_2 \models A' \sqsubseteq \neg A'$.

In general, it is not possible to have a target TBox $\mathcal{T}_2$ such that (A, D) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent and $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, that is, it is not possible "contradict" concepts A and D in the target. ■

We explained in the examples above that in order to check whether $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$ one needs to verify whether the axioms implied by $\mathcal{T}_1 \cup \mathcal{T}_{12}$ are "representable", and whether $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent pairs are "target contradictable". To define these notions required for the characterization in Lemma 7.2.8, we first introduce the notion of reserved UCQ-representation. We say that a target TBox $\mathcal{T}_2$ is a *reserved* UCQ-*representation* of $\mathcal{T}_1$ under $\mathcal{M}$, if for every ABox $\mathcal{A}_1$ over Σ_1 that is consistent with $\mathcal{T}_1$, $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ Σ_2-query entails $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$. Observe that the empty TBox is a reserved UCQ-representation.

In the definitions below, X and Y denote basic concepts or roles over Σ_1, and X' denotes a basic concept or role over Σ_2.

Definition 7.2.3. *Inclusion $X \sqsubseteq X'$ is* representable *in $\mathcal{T}_1$ and $\mathcal{M}$, if there exists a target axiom α (possibly trivial) such that whenever $\mathcal{T}_2$ is a reserved* UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$, it holds that $\mathcal{T}_2' = \mathcal{T}_2 \cup \{\alpha\}$ is also a reserved* UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$, moreover $\mathcal{T}_2' \cup \mathcal{T}_{12} \models X \sqsubseteq X'$.*

In that case, $X \sqsubseteq X'$ is representable by *α.*

Definition 7.2.4. *Pair (X, Y) is* target contradictable *in $\mathcal{T}_1$ and $\mathcal{M}$, if there exists a target axiom α (possibly trivial) such that whenever $\mathcal{T}_2$ is a reserved* UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$, it holds that $\mathcal{T}_2' = \mathcal{T}_2 \cup \{\alpha\}$ is also a reserved* UCQ-*representation of $\mathcal{T}_1$ under $\mathcal{M}$, moreover (X, Y) is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent.*

In that case, (X, Y) is target contradictable by *α.*

Our last definition before we present a characterization of the cases when $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$ is the notion of a *generating pass*. In the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$ for some concept B and role R, existence of a generating pass for (B, R) ensures that there exists a reserved UCQ-representation $\mathcal{T}_2$ satisfying condition (iii) for B and R. The main reason behind this non-trivial definition is to ensure that for R that is not translated via the mapping, i.e., $\mathcal{T}_1 \cup \mathcal{T}_{12} \not\models R \sqsubseteq R'$ for any role R' over Σ_2, there exists an "image" $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ that need not be a neighbor of o and that is already a "one-to-one" image of some other $x \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$. For instance, in Example 7.1.5, for $B = A$, this y was equal to $w_{Q'}$, which was a "one-to-one image" of $w_Q \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}$. To the contrast, if we find $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ that is not a "one-to-one image" of any element in $\Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$, then such $\mathcal{T}_2$ could not be a UCQ-representation, as demonstrated in Example 7.1.4.

For a TBox $\mathcal{T}$ and a concept B, denote by $\mathbf{u}_\Sigma^\mathcal{T}(B)$ the set of all concepts B' over Σ such that $\mathcal{T} \models B \sqsubseteq B'$; $\mathbf{u}_\Sigma^\mathcal{T}(R)$ is defined analogously for a role R.

Definition 7.2.5. *Let B be a concept over Σ_1 and R a role. A* generating pass for (B, R) *in $\mathcal{T}_1$ and $\mathcal{M}$ is a tuple of concepts $\langle C_0, C_1, \ldots C_n \rangle$ of length greater or equal 1, such that $C_0 = B$, and it holds for $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_{12}$ and $i = 0, \ldots, n-1$*

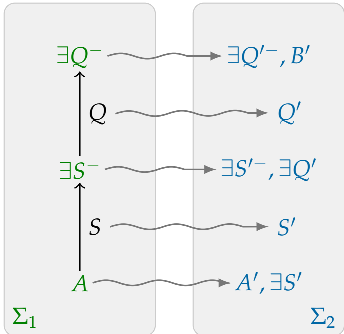
(RCHAIN) $C_{i+1} = \exists Q_i^-$ *for some role* Q_i *s.t.* $\mathcal{T} \models C_i \sqsubseteq \exists Q_i$ *and* $\mathbf{u}_{\Sigma_2}^{\mathcal{T}}(Q_i) \neq \emptyset$;

(CPASS) *for each* $D_i \in \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_i)$, *inclusion* $C_i \sqsubseteq D_i$ *is representable in* $\mathcal{T}_1$ *and* $\mathcal{M}$;

(RPASS) *for each* $S_i \in \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(Q_i)$, *inclusion* $Q_i \sqsubseteq S_i$ *is representable in* $\mathcal{T}_1$ *and* $\mathcal{M}$;

(ETYPE) $\mathbf{u}_{\Sigma_2}^{\mathcal{T}}(\exists R^-) \subseteq \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_n)$, *and* $\mathbf{u}_{\Sigma_2}^{\mathcal{T}}(R) \subseteq \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_0, C_n)$.

Example 7.2.6. Assume $\mathcal{M}$ and $\mathcal{T}_1$ from Example 7.1.5. Then $\langle A, \exists S^-, \exists Q^- \rangle$ is a generating pass for (A, R) in $\mathcal{T}_1$ and $\mathcal{M}$. Below we represent it graphically, where the Σ_2 super concepts and super roles of concepts and roles over Σ_1 are shown at the ends of the "wavy" arrows.



■

Example 7.2.7. Assume $\mathcal{M}$ and $\mathcal{T}_1$ from Example 7.1.4. Then there exists no generating pass for (A, R) in $\mathcal{T}_1$ and $\mathcal{M}$. ■

We also make use of the following properties.

(TINCONSC) concept B is $\mathcal{T}$-inconsistent iff $\mathcal{T} \models B \sqsubseteq C \sqcap D$ for some concept disjointness $C \sqsubseteq \neg D \in \mathcal{T}$, or there exist $n \geq 1$ and roles $R_1, \ldots, R_n$ such that $\mathcal{T} \models \{B \sqsubseteq_{\rightsquigarrow} \exists R_1, \exists R_i^- \sqsubseteq_{\rightsquigarrow} \exists R_{i+1}\}$, and

- $\mathcal{T} \models \exists R_n^- \sqsubseteq C \sqcap D$, for some concept disjointness $C \sqsubseteq \neg D \in \mathcal{T}$, or
- $\mathcal{T} \models R_n \sqsubseteq S \sqcap Q$ or $\mathcal{T} \models R_n \sqsubseteq S^- \sqcap Q^-$, for some role disjointness $S \sqsubseteq \neg Q \in \mathcal{T}$.

(TINCONSR) role R is $\mathcal{T}$-inconsistent iff $\mathcal{T} \models R \sqsubseteq S \sqcap Q$ or $\mathcal{T} \models R \sqsubseteq S^- \sqcap Q^-$ for some role disjointness $S \sqsubseteq \neg Q \in \mathcal{T}$, or one of $\exists R, \exists R^-$ is $\mathcal{T}$-inconsistent.

Having defined all notions above, we provide a characterization of the cases when $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$, which has a similar structure to the characterization of UCQ-representations in Lemma 7.1.1.

Lemma 7.2.8. *Given a mapping* $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ *and a TBox* $\mathcal{T}_1$ *over* Σ_1, $\mathcal{T}_1$ *is* UCQ-*represen-table under* $\mathcal{M}$, *if and only if the following conditions are satisfied:*

(I) *For each* $\mathcal{T}_1$*-consistent pair of concepts or roles* X, Y *over* Σ_1 *such that* (X, Y) *is* $\mathcal{T}_1 \cup \mathcal{T}_{12}$*-inconsistent,* (X, Y) *is target contradictable in* $\mathcal{T}_1$ *and* $\mathcal{M}$.

(II) *For each* $\mathcal{T}_1 \cup \mathcal{T}_{12}$*-consistent concept or role* X *over* Σ_1 *and each* X' *over* Σ_2 *such that* $\mathcal{T}_1 \cup \mathcal{T}_{12} \models X \sqsubseteq X'$, *inclusion* $X \sqsubseteq X'$ *is representable in* $\mathcal{T}_1$ *and* $\mathcal{M}$.

(III) *For each* $\mathcal{T}_1 \cup \mathcal{T}_{12}$*-consistent concept* B *over* Σ_1 *and each role* R *such that* $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$, *there exists a generating pass for* (B, R) *in* $\mathcal{T}_1$ *and* $\mathcal{M}$.

Proof. ($\Leftarrow$) Assume conditions (I) – (III) are satisfied, we construct a TBox $\mathcal{T}_2$ over Σ_2 and prove it is a UCQ-representation for $\mathcal{T}_1$ under $\mathcal{M}$. The required $\mathcal{T}_2$ will be given as the union of the three sets of axioms presented below. First, let (B, C) be a $\mathcal{T}_1$-consistent and $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent pair of concepts over Σ_1, then (B, C) is target contradictable by condition (I): assume (B, C) is target contradictable by α, then define set $ax_i(B, C)$ to be equal to $\{\alpha\}$. Similarly, we define $ax_i(R, Q) = \{\alpha\}$ for $\mathcal{T}_1$-consistent and $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent pair of roles R, Q over Σ_1. Next, take a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1, and assume $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq C'$ for C' over Σ_2, then by condition (II), $B \sqsubseteq C'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$: let $ax_{ii}(B, C') = \{\alpha\}$ such that $B \sqsubseteq C'$ ia representable by α. Similarly, for a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent role R over Σ_1 and Q' over Σ_2, such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models R \sqsubseteq Q'$. Finally, for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1 and each role R such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R$, define the set $ax_{iii}(B, R)$ from the generating pass $\langle C_0, \dots, C_n \rangle$ for (B, R) in $\mathcal{T}_1$ and $\mathcal{M}$ given by condition (III). Take $ax_{iii}(B, R)$ equal to the set of all axioms α, where $C_i \sqsubseteq D_i$ is representable by α in (cpass), or $Q_i \sqsubseteq S_i$ is representable by α in (rpass). Finally we have:

$$\mathcal{T}_2 = \bigcup_{\substack{X,Y \text{ conc. or roles over } \Sigma_1,\\ \mathcal{T}_1\text{-consistent and}\\ \mathcal{T}_1 \cup \mathcal{T}_{12}\text{-inconsistent}}} ax_i(X, Y) \;\cup \bigcup_{\substack{X \text{ conc. or role over } \Sigma_1,\\ \mathcal{T}_1 \cup \mathcal{T}_{12}\text{-consistent },\\ X' \text{ over } \Sigma_2,\, \mathcal{T}_1 \cup \mathcal{T}_{12} \models X \sqsubseteq X'}} ax_{ii}(X, X') \;\cup \bigcup_{\substack{\mathcal{T}_1 \cup \mathcal{T}_{12}\text{-cons. } B \text{ over } \Sigma_1,\\ \mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\rightsquigarrow} \exists R}} ax_{iii}(B, R)$$

Then it immediately follows that $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$: On the one hand, by construction, $\mathcal{T}_2$ is a reserved UCQ-representation. On the other hand, the $\Rightarrow$ directions of conditions (i) and (ii), and condition (iii) of Lemma 7.1.1 are satisfied by construction of $\mathcal{T}_2$ and by definition of ax_i, ax_{ii}, ax_{iii}. From which, it follows that for each ABox $\mathcal{A}_1$ consistent with $\mathcal{T}_1$, $\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$ Σ_2-query entails $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$. Hence, indeed, $\mathcal{T}_2$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$.

Below we show satisfaction of condition (iii) of Lemma 7.1.1. Denote by $\mathcal{T}$ the TBox $\mathcal{T}_1 \cup \mathcal{T}_{12}$, let B be $\mathcal{T}$-consistent concept over Σ_1 and R a role $\mathcal{T} \models B \sqsubseteq_{\rightsquigarrow} \exists R$, moreover let $\langle C_0, \dots, C_n \rangle$ be a generating pass for (B, R) in $\mathcal{T}_1$ and $\mathcal{M}$ given by condition (III) such that $n \geq 0$, $C_0 = B$, $C_{i+1} = \exists Q_i^-$ and

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(w_{[R]}) \subseteq \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_n) \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(o, w_{[R]}) \subseteq \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_0, C_n).$$

For $n = 0$, if $\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(w_{[R]}) = \emptyset$, then we take $y \in \Delta^{\mathcal{G}\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\} \rangle}$ equal to o. If $\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(w_{[R]}) \neq \emptyset$, then $\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(o)$ and again, we take $y = o$. If $n > 0$, it follows by (rpass) that $\mathbf{u}_{\Sigma_2}^{\mathcal{T}}(Q_i) \neq \emptyset$, and $\mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_i) \neq \emptyset$. Consequently, by construction of $ax_{iii}(B, R)$, $\mathcal{T}_2 \cup \mathcal{T}_{12} \models C_i \sqsubseteq D_i$ for each $D_i \in \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(C_i)$, and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models Q_i \sqsubseteq S_i$ for each $S_i \in \mathbf{u}_{\Sigma_2}^{\mathcal{T}}(Q_i)$. Recall the shape of $\mathcal{T}_{12}$ and $\mathcal{T}_2$: it follows there exists $y \in \Delta^{\mathcal{G}\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\} \rangle}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\} \rangle}(y) \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}, \{B(o)\} \rangle}(o, w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\} \rangle}(o, y).$$

($\Rightarrow$) Let $\mathcal{T}_2$ be a UCQ-representation for $\mathcal{T}_1$ under $\mathcal{M}$. It is easy to see that conditions (I) and (II) are satisfied.

We show (III) is satisfied; Denote by $\mathcal{T}$ the TBox $\mathcal{T}_1 \cup \mathcal{T}_{12}$, assume B is a $\mathcal{T}$-consistent concept over Σ_1 and $\mathcal{T} \models B \underset{\rightsquigarrow}{\sqsubseteq} \exists R$ for some role R, by condition (iii) of Lemma 7.1.1 it follows there exists $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(y), \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle}}(o, y).$$

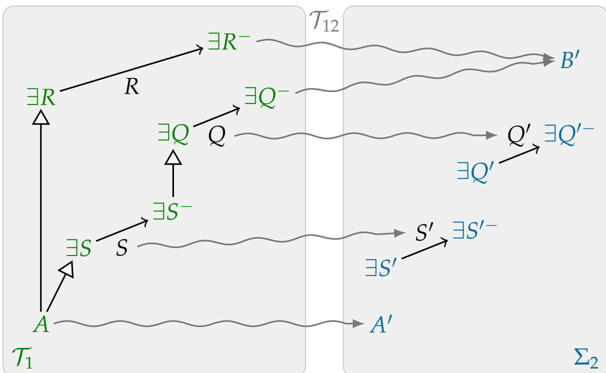
Assume that $y = w_{[Q_n]}$ for $n \geq 0$, where $o \rightsquigarrow_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B(o)\}\rangle} w_{[Q_1]} \rightsquigarrow \cdots \rightsquigarrow w_{[Q_n]}$. Then $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \{B \sqsubseteq \exists Q_1, \exists Q_i^- \sqsubseteq \exists Q_{i+1}, \exists Q_n^- \sqsubseteq B'\}$, for all $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(w_{[R]})$, and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \{Q_1 \sqsubseteq R' \mid R' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(o, w_{[R]})\}$ if $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(o, w_{[R]}) \neq \emptyset$. One can show by induction that for each i, there exist S_i over Σ_1 such that $\mathcal{T} \models S_i \sqsubseteq Q_i$ and $\mathcal{T} \models \{B \sqsubseteq \exists S_i, \exists S_i^- \sqsubseteq \exists S_{i+1}, \exists S_n^- \sqsubseteq B'\}$, for all $B' \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}, \{B(o)\}\rangle}}(w_{[R]})$. We define a sequence $\langle C_0, \ldots, C_n \rangle$ as $C_0 = B$, $C_{i+1} = \exists S_i^-$: it can be straightforwardly verified that $\langle C_0, \ldots, C_n \rangle$ is a generating pass for (B, R) in $\mathcal{T}_1$ and $\mathcal{M}$. □

We now use the above characterization to verify UCQ-representability in the following examples.

Example 7.2.9. Assume $\mathcal{M}$ and $\mathcal{T}_1$ from Example 7.1.5, that is, $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where

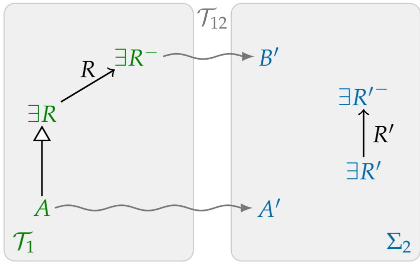
$$\begin{aligned} \Sigma_1 &= \{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot), Q(\cdot,\cdot)\} \\ \Sigma_2 &= \{A'(\cdot), B'(\cdot), S'(\cdot,\cdot), Q'(\cdot,\cdot)\} \\ \mathcal{T}_{12} &= \{A \sqsubseteq A', \exists R^- \sqsubseteq B', \\ &\qquad S \sqsubseteq S', Q \sqsubseteq Q', \exists Q^- \sqsubseteq B'\} \end{aligned} \quad \text{and} \quad \begin{aligned} \mathcal{T}_1 = \{&A \sqsubseteq \exists R, \\ &A \sqsubseteq \exists S, \exists S^- \sqsubseteq \exists Q\} \end{aligned}$$

Then the ER diagrams of $\mathcal{T}_1$ and $\mathcal{T}_{12}$ can be depicted as follows:



Then one can see that conditions (I) – (III) are satisfied. Thus, for instance, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq \exists S'$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists S^- \sqsubseteq \exists Q'$: clearly both inclusions are representable in $\mathcal{T}_1$ and $\mathcal{M}$. Then, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \underset{\rightsquigarrow}{\sqsubseteq} \exists R$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \underset{\rightsquigarrow}{\sqsubseteq} \exists S$, and in both cases there exist generating passes: $\langle A, \exists S^-, \exists Q^- \rangle$ from Example 7.2.6 and $\langle A, \exists S^- \rangle$, respectively. This confirms that $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$. ■

Example 7.2.10. Assume $\mathcal{M}$ and $\mathcal{T}_1$ from Example 7.1.4, that is, $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot)\}$, $\Sigma_2 = \{A'(\cdot), R'(\cdot,\cdot), B'(\cdot)\}$, $\mathcal{T}_{12} = \{A \sqsubseteq A', \exists R^- \sqsubseteq B'\}$, and $\mathcal{T}_1 = \{A \sqsubseteq \exists R\}$. The ER diagrams of $\mathcal{T}_1$ and $\mathcal{T}_{12}$ can be depicted as follows:



In contrast with the previous example, condition (III) is not satisfied. In fact, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq \exists R$, however there exists no generating pass $\langle C_0, \ldots, C_n \rangle$ for (A, R) in $\mathcal{T}_1$ and $\mathcal{M}$, such that $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{A(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{u}_{\Sigma_2}^{\mathcal{T}_1 \cup \mathcal{T}_{12}}(C_n)$. So indeed, $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$. ∎

Example 7.2.11. Assume that $\mathcal{M} = (\{A(\cdot), B(\cdot), C(\cdot)\}, \{A'(\cdot), B'(\cdot)\}, \mathcal{T}_{12})$, where $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq \neg A'\}$, and let $\mathcal{T}_1 = \{B \sqsubseteq C\}$. (This case was considered in Example 4.3.3–(3).) We show condition (I) is satisfied: the pairs (A, C) and (A, B) are $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. As the former pair is already $\mathcal{T}_{12}$-inconsistent, this case is not interesting. Consider the latter pair, then one can easily verify that (A, B) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$: $A \sqsubseteq A' \in \mathcal{T}_{12}$ and disjointness $B \sqsubseteq \neg A'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ by $B' \sqsubseteq \neg A'$, as only A is transfered positively to A', only B is transfered positively to B', and it is not the case that (A, B) is a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent pair. It is easy to see that condition (II) is satisfied, because for concepts D and D' over Σ_1 and over Σ_2, respectively, it holds that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq D'$ implies $\mathcal{T}_{12} \models D \sqsubseteq D'$. ∎

Finally, we obtain the complexity bound of the non-emptiness problem for UCQ-representations.

Theorem 7.2.12. *The non-emptiness problem for* UCQ*-representations is* NLOGSPACE*-complete.*

Proof. As in the case of Theorem 7.1.8, the lower bound is shown by the reduction from the directed graph reachability problem, however, we need a slightly more involved encoding.To encode the graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$, we need a set of Σ_1-concept names $\{V_i \mid v_i \in \mathsf{V}\} \cup \{S, F, X, Y\}$ and a set of Σ_2-concept names $\{V_i' \mid v_i \in \mathsf{V}\} \cup \{S', X', Y'\}$. Consider the TBox

$$\mathcal{T}_1 = \{V_i \sqsubseteq V_j \mid (v_i, v_j) \in \mathsf{E}\} \cup \{S \sqsubseteq V_k, V_m \sqsubseteq F, X \sqsubseteq Y\},$$

where v_k and v_m are, respectively, the initial and final vertices. Then, let

$$\mathcal{T}_{12} = \{V_i \sqsubseteq V_i' \mid v_i \in \mathsf{V}\} \cup \{S \sqsubseteq S', S \sqsubseteq X', F \sqsubseteq Y', X \sqsubseteq X', Y \sqsubseteq Y'\};$$

we will show:

- there is a directed path from v_k to v_m in G iff there exists a UCQ-representation for $\mathcal{T}_1$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$.

Indeed, using Lemma 7.2.8, there exists a representation iff condition (II) is satisfied. By the structure of $\mathcal{T}_1 \cup \mathcal{T}_{12}$ one can see that it is the case iff inclusions $X \sqsubseteq Y'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ by $X' \sqsubseteq Y'$, i.e., iff $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq X'$ implies $\mathcal{T}_1 \cup \mathcal{T}_{12} \models$

$S \sqsubseteq Y'$, and that holds iff $\mathcal{T}_1 \models S \sqsubseteq F$, which is the case iff there exists a path from v_k to v_m in G. This completes the proof of the lower bound.

To show the upper bound, we prove that conditions (I) – (III) of Lemma 7.2.8 can be checked in NLogSpace. First, there are syntactic conditions that allow to check whether an inclusion is representable in $\mathcal{T}_1$ and $\mathcal{M}$, and whether a pair is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$ (see Propositions 7.2.13, 7.2.14, 7.2.15 and 7.2.16). In fact, these conditions can be checked using the algorithm, based on directed graph reachability solving procedure, similar to the proof of Theorem 7.1.8. The only new case is condition (III); to verify for a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1 and a role R such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq_{\leadsto} \exists R$, that there exists a generating pass $\pi = \langle C_0, \ldots C_n \rangle$ for (B, R) in $\mathcal{T}_1$ and $\mathcal{M}$, we can use the following procedure, running in NLogSpace. First, we take $C_0 = B$ and decide, if the pass ends here (i.e., $n = 0$). If we decided so, it only remains verify (cpass). This verification can be performed in NLogSpace, similarly to the method described in the proof of Theorem 7.1.8. If, on the other hand, we decide, that the pass continues, we "guess" $C_1 = \exists Q^-$ for some role Q, and verify that (rchain) and (cpass) are satisfied. Now, if we decide that the pass stops, it remains to verify (rpass). If, on the contrary, we decide that the pass continues, we can "forget" C_0, "guess" C_2, and proceed with it in the same way, as we did with C_1. Finally, when we reach the concept C_n, such that the algorithm decides to stop, it remains to verify (cpass). It should be clear that whenever a generating pass $\pi = \langle C_0, \ldots C_n \rangle$ for (B, R) in $\mathcal{T}_1$ and $\mathcal{M}$ exists, we can find it by the above non-determinictic procedure.

Below, we provide propositions that establish the syntactic conditions for checking whether an inclusion is representable and whether a pair is target contradictable.

Proposition 7.2.13. *For a concept B over Σ_1 and C' over Σ_2, inclusion $B \sqsubseteq C'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ if and only if there exists B' over Σ_2 such that $\mathcal{T}_{12} \models B \sqsubseteq B'$, and for each $\mathcal{T}_1$-consistent concept D over Σ_1:*

(CINCL) *$\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq B'$ implies $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq C'$,*

(MROLE) *if $B' = \exists Q'^-$ for some role Q' over Σ_2, then $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \exists Q'$ implies $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\leadsto} \exists Q$ for some role Q s.t. $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{Q \sqsubseteq Q', \exists Q^- \sqsubseteq C'\}$.*

In this case, $B \sqsubseteq C'$ is representable by *$B' \sqsubseteq C'$.*

Proof. ($\Leftarrow$) Let B be a concept over Σ_1 and C' over Σ_2, $B' \neq C'$, and conditions (cincl) and (mrole) are satisfied. We show inclusion $B \sqsubseteq C'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ by $B' \sqsubseteq C'$. Take $\mathcal{T}_2$ a reserved UCQ-representation for $\mathcal{T}_1$ under $\mathcal{M}$: we prove $\mathcal{T}_2' = \mathcal{T}_2 \cup \{B' \sqsubseteq C'\}$ is a reserved UCQ-representation by showing the following is satisfied:

- for each $\mathcal{T}_1$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent pair of concepts or roles (X, Y), it follows (X, Y) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which corresponds to the $\Leftarrow$ direction of condition (i) of Lemma 7.1.1,

- for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept or role X over Σ_1 and each X' over Σ_2, $\mathcal{T}_2' \cup \mathcal{T}_{12} \models X \sqsubseteq X'$ implies $\mathcal{T}_1 \cup \mathcal{T}_{12} \models X \sqsubseteq X'$, which corresponds to the $\Leftarrow$ direction of condition (ii) of Lemma 7.1.1, and

- condition (iv) of Lemma 7.1.1.

Observe that from $\mathcal{T}_2$ is a reserved UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}$, it follows the above conditions are already satisfied for $\mathcal{T}_2$, $\mathcal{T}_1$ and $\mathcal{M}$.

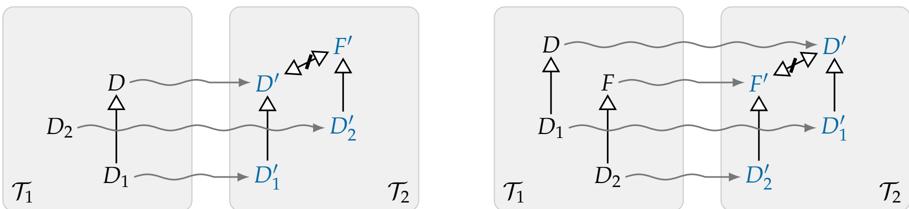
First, for condition (ii) of Lemma 7.1.1, let D be a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept over Σ_1 and E' a concept over Σ_2 such that $\mathcal{T}_2' \cup \mathcal{T}_{12} \models D \sqsubseteq E'$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models D \sqsubseteq E'$. Hence, there exists D' over Σ_2 such that $\mathcal{T}_2 \models \{D' \sqsubseteq B', C' \sqsubseteq E'\}$ and $\mathcal{T}_{12} \models D \sqsubseteq D'$. Since $\mathcal{T}_2$ is a reserved UCQ-representation and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D \sqsubseteq B'$, it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq B'$, so there exists B_1 over Σ_1 such that $\mathcal{T}_1 \models D \sqsubseteq B_1$ and $\mathcal{T}_{12} \models B_1 \sqsubseteq B'$. Next, B', C' satisfy condition (cincl), therefore $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq C'$, so there exists C over Σ_1 such that $\mathcal{T}_1 \models B_1 \sqsubseteq C$ and $\mathcal{T}_{12} \models C \sqsubseteq C'$. And we can continue by analogy. To summarize, there exist B_1, C and E over Σ_1 such that

$$\mathcal{T}_1 \models \{D \sqsubseteq B_1, B_1 \sqsubseteq C, C \sqsubseteq E\} \tag{16}$$

and $\mathcal{T}_{12} \models \{B_1 \sqsubseteq B', C \sqsubseteq C', E \sqsubseteq E'\}$. Finally, we obtain that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq E'$.

Next, for condition (i), let (D_1, D_2) be a pair of $\mathcal{T}_1$-consistent, $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent concepts. For the sake of contradiction, assume (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent (hence, each D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent).

Suppose both D_i are $\mathcal{T}_2' \cup \mathcal{T}_{12}$-consistent. Without loss of generality, we may assume that for some D' over Σ_2, $\mathcal{T}_2' \cup \mathcal{T}_{12} \models \{D_1 \sqsubseteq D',\ D_2 \sqsubseteq \neg D'\}$. From condition ii, it follows there exists D over Σ_1 such that $\mathcal{T}_1 \models D_1 \sqsubseteq D$ and $\mathcal{T}_{12} \models D \sqsubseteq D'$. Consider the following cases:
1) $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D_2 \sqsubseteq \neg D'$ (and $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models D_1 \sqsubseteq D'$). Then, either there exist D_2', F' over Σ_2 such that $\mathcal{T}_2 \models \{D_2' \sqsubseteq F', F' \sqsubseteq \neg D'\}$ and $\mathcal{T}_{12} \models D_2 \sqsubseteq D_2'$ (see the diagram below), or $\mathcal{T}_{12} \models D_2 \sqsubseteq \neg D'$. In both cases, (D, D_2) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, so it follows (D, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. In view of $\mathcal{T}_1 \models D_1 \sqsubseteq D$, we obtain contradiction with the assumption (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent.
2) $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models D_2 \sqsubseteq \neg D'$. Then, there exists F' over Σ_2 such that $\mathcal{T}_2' \cup \mathcal{T}_{12} \models D_2 \sqsubseteq F'$ and $\mathcal{T}_2 \models F' \sqsubseteq \neg D'$ (note, $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models D_2 \sqsubseteq F'$). From condition ii, it follows there exists F over Σ_1 such that $\mathcal{T}_1 \models D_2 \sqsubseteq F$ and $\mathcal{T}_{12} \models F \sqsubseteq F'$. Now, as (D, F) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, it follows (D, F) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which in view of $\mathcal{T}_1 \models \{D_1 \sqsubseteq D, D_2 \sqsubseteq F\}$ contradicts the assumption (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent.



(1) (2)

Suppose one of D_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent. Consider the following two cases by (tinconsc):
1) for some D' over Σ_2, $\mathcal{T}_2' \cup \mathcal{T}_{12} \models \{D_i \sqsubseteq D',\ D_i \sqsubseteq \neg D'\}$. The contradiction is obtained similarly as in the case both D_i are $\mathcal{T}_2' \cup \mathcal{T}_{12}$-consistent.
2) there exist $n \geq 1$ and distinct roles $S_1', \ldots, S_n'$ such that $\mathcal{T}_2' \cup \mathcal{T}_{12} \models \{D_i \mathrel{\underset{\rightsquigarrow}{\sqsubseteq}} \exists S_1',\ \exists S_j'^{-} \mathrel{\underset{\rightsquigarrow}{\sqsubseteq}} \exists S_{j+1}'\}$ and $\mathcal{T}_2' \cup \mathcal{T}_{12} \models S_n' \sqsubseteq R' \sqcap Q'$ for $R' \sqsubseteq \neg Q' \in \mathcal{T}_2$, or $\mathcal{T}_2' \cup \mathcal{T}_{12} \models \exists S_n'^{-} \sqsubseteq E' \sqcap F'$ for $E' \sqsubseteq \neg F' \in \mathcal{T}_2$.

If $n = 1$ and S_1' is a role over Σ_1 (i.e., $D_i = \exists S_1'$ and S_1' is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent), then from condition ii, it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S_1' \sqsubseteq R' \sqcap Q'$ or $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists S_1'^{-} \sqsubseteq E' \sqcap F'$.

In the former case, there exist roles R, Q over Σ_1 such that $\mathcal{T}_1 \models S'_1 \sqsubseteq R \sqcap Q$ and $\mathcal{T}_{12} \models \{R \sqsubseteq R',\ Q \sqsubseteq Q'\}$; then (R, Q) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, since $\mathcal{T}_2$ is a reserved UCQ-representation, it follows (R, Q) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. In the latter case, there exist concepts E, F over Σ_1 such that $\mathcal{T}_1 \models \exists S'^{-}_1 \sqsubseteq E \sqcap F$ and $\mathcal{T}_{12} \models \{E \sqsubseteq E',\ F \sqsubseteq F'\}$; then (E, F) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, hence (E, F) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. In any case we obtain S'_1 is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent.

If $n = 1$ and S'_1 is a role over Σ_2, assume $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models D_i \sqsubseteq \exists S'_1$. From condition (ii) it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D_i \sqsubseteq \exists S'_1$, so there exists D over Σ_1 such that $\mathcal{T}_1 \models D_i \sqsubseteq D$ and $\mathcal{T}_{12} \models D \sqsubseteq \exists S'_1$. Then $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists T'$ for some role T' (possibly coinciding with S'_1) such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models T' \sqsubseteq S'_1$. In the case $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S'_1 \sqsubseteq R' \sqcap Q'$ or $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \exists S'^{-}_1 \sqsubseteq E' \sqcap F'$, since $\mathcal{T}_2$ is a reserved UCQ-representation, from condition (iv) it follows there exists a role T such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists T$, and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models T \sqsubseteq R' \sqcap Q'$ or $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists T^{-} \sqsubseteq E' \sqcap F'$. Again, we obtain that D is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent.

Assume now $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models \exists S'^{-}_1 \sqsubseteq E' \sqcap F'$ (the case $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models S'_1 \sqsubseteq R' \sqcap Q'$ is not possible). Then it follows $\mathcal{T}_2 \models \{\exists S'^{-}_1 \sqsubseteq B',\ C' \sqsubseteq E'\}$ and/or $\mathcal{T}_2 \models \{\exists S'^{-}_1 \sqsubseteq B',\ C' \sqsubseteq F'\}$, and the role T above is such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists T^{-} \sqsubseteq B'$. If T is over Σ_1, then $\mathcal{T}_1 \models \exists T^{-} \sqsubseteq B_1$ and $\mathcal{T}_{12} \models B_1 \sqsubseteq B'$ for some concept B_1 over Σ_1, next we have that $\mathcal{T}'_2 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq E' \sqcap F'$, so from condition (ii) it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq E' \sqcap F'$, and as before B_1 is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. If T is over Σ_2, then $B' = \exists T^{-} = \exists S^{-}_1$, and by (mrole) it follows there exists S_1 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists S_1$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S_1 \sqsubseteq S'_1,\ \exists S^{-}_1 \sqsubseteq C'\}$. Since $\exists S'^{-}_1 \neq C'$, it follows S_1 is over Σ_1, and there exists C over Σ_1 such that $\mathcal{T}_1 \models \exists S^{-}_1 \sqsubseteq C$ and $\mathcal{T}_{12} \models C \sqsubseteq C'$. Now, we have that $\mathcal{T}'_2 \cup \mathcal{T}_{12} \models C \sqsubseteq E' \sqcap F'$, from condition (ii) it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models C \sqsubseteq E' \sqcap F'$, so as before C is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent.

For $n > 1$, we can continue reasoning as for the case $n = 1$ to obtain a contradiction. Finally, we conclude that D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, hence (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent.

Let (S_1, S_2) be a pair of $\mathcal{T}_1$-consistent, $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}'_2 \cup \mathcal{T}_{12}$-inconsistent roles (this is the only non-trivial case). Since $\mathcal{T}'_2$ extends $\mathcal{T}_2$ with a concept inclusion, we have that there exist D_1, D_2 covering $\{\exists S_1, \exists S_2\}$ or $\{\exists S^{-}_1, \exists S^{-}_2\}$ such (D_1, D_2) is $\mathcal{T}'_2 \cup \mathcal{T}_{12}$-inconsistent and $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent. By reasoning as above, we obtain (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, therefore (S_1, S_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent.

To show condition (iv) of Lemma 7.1.1 assume a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept D over Σ_1 and a role R such that $\mathcal{T}'_2 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists R$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \not\models D \sqsubseteq_{\rightsquigarrow} \exists R$. Hence, R is a role over Σ_2, and there exists D' over Σ_2 such that $\mathcal{T}_2 \models \{D' \sqsubseteq B', C' \sqsubseteq \exists R\}$ and $\mathcal{T}_{12} \models D \sqsubseteq D'$. As before, we can conclude there exists (a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent) C over Σ_1 such that $\mathcal{T}_{12} \models C \sqsubseteq C'$ (and $\mathcal{T}_1 \models D \sqsubseteq C$). It means $\mathcal{T}_2 \cup \mathcal{T}_{12} \models C \sqsubseteq \exists R$, therefore either $\mathcal{T}_2 \cup \mathcal{T}_{12} \models C \sqsubseteq_{\rightsquigarrow} \exists R$, or $C = \exists Q$ for some role Q over Σ_1 such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models Q \sqsubseteq R$, and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models C \sqsubseteq_{\rightsquigarrow} \exists Q$. Since $\mathcal{T}_2$ is a reserved UCQ-representation, it follows there exists $z \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{C(o)\}\rangle}}$ such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{C(o)\}\rangle}}(x) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{C(o)\}\rangle}}(z) \quad \text{and} \quad \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{C(o)\}\rangle}}(o, x) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{C(o)\}\rangle}}(o, z),$$

with $x = w_{[R]}$ or $x = w_{[Q]}$. Observe that $R \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{C(o)\}\rangle}}(o, x)$, which implies that $z = w_{[S]}$ for some role S such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models C \sqsubseteq \exists S$. Now, notice that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models$

$D \sqsubseteq \exists S$: we obtain that $o \rightsquigarrow_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle} w_{[T]}$ for some role T (possibly coinciding with S) such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models T \sqsubseteq S$. Finally, we have that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[T]}),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(o, w_{[R]}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(o, w_{[T]}),$$

so we take y in condition (ii) to be equal to $w_{[T]}$.

Assume now $B' = \exists R^-$ for some role R over Σ_2, and D is a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept over Σ_1 such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists R$. By condition (ii), it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \exists R$. The interesting case to consider is $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[R]}) = \{\exists R^-\}$ (hence, $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(o, w_{[R]}) = \{R\}$), as for $\mathcal{T}_2$ it is enough to take $y \in \Delta^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}$ equal to $w_{[S]}$ such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists S$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R$ (such S exists: we take S equal to R if $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists R$). However, given the axiom $\exists R^- \sqsubseteq C'$ in $\mathcal{T}_2'$, we have $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[R]}) \supseteq \{\exists R^-, C'\}$ (note, still $\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(o, w_{[R]}) = \{R\}$). As B' and C' satisfy (mrole) and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \exists R$, it follows there exists S such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists S$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S \sqsubseteq R,\ \exists S^- \sqsubseteq C'\}$; moreover by $C' \neq \exists R^-$ and the structure of $\mathcal{T}_1 \cup \mathcal{T}_{12}$ it follows S is over Σ_1. From the latter we obtain a role Q over Σ_1 such that $\mathcal{T}_1 \models S \sqsubseteq Q$ and $\mathcal{T}_{12} \models Q \sqsubseteq R$, moreover $\exists Q^-$ and Q are $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Now, assume $\mathcal{T}_2 \models \exists R^- \sqsubseteq E'$; then $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \exists Q^- \sqsubseteq E'$, and since $\mathcal{T}_2$ satisfies condition (ii) it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists Q^- \sqsubseteq E'$, therefore $E' \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[S]})$. Thus $\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[R]}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}}(w_{[S]})$, and we take $y = w_{[S]}$ to satisfy condition (iv) of Lemma 7.1.1.

($\Rightarrow$) Suppose inclusion $B \sqsubseteq C'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ by a target axiom α. Then $\mathcal{T}_2 = \{\alpha\}$ is a reserved UCQ-representation and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B \sqsubseteq C'$. If $\mathcal{T}_{12} \models B \sqsubseteq C'$, we take B' equal to C': obviously, (cincl) and (mrole) are satisfied. Now, assume $\mathcal{T}_{12} \not\models B \sqsubseteq C'$. Then it must be the case α is of the form $D' \sqsubseteq C'$ and $\mathcal{T}_{12} \models B \sqsubseteq D'$ for some concept D' over Σ_2. So we take B' equal to D', and prove below (cincl) and (mrole) are satisfied.

For (cincl), let $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq B'$ for a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept D over Σ_1. It follows $\mathcal{T}_1 \models D \sqsubseteq B_1$ and $\mathcal{T}_{12} \models B_1 \sqsubseteq B'$ for some concept B_1 over Σ_1. Consequently, $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq C'$, and as $\mathcal{T}_2$ is a reserved UCQ-representation, we obtain that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq C'$. Finally, we proved that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq C'$.

For (mrole), assume B' is of the form $\exists Q'^-$ for some role Q' over Σ_2, and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \exists Q'$. As above, there exists B_1 over Σ_1 such that $\mathcal{T}_{12} \models B_1 \sqsubseteq \exists Q'$. Then, $\mathcal{T}_2 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq_{\rightsquigarrow} \exists S'$ for some role S' (possibly coinciding with Q') such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S' \sqsubseteq Q'$. By condition (iv) of Lemma 7.1.1 and $Q' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B_1(o)\}\rangle}}(o, w_{[S']})$, there exists a role S such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B_1(o)\}\rangle}}(w_{[S']}) \subseteq \mathbf{t}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B_1(o)\}\rangle}}(w_{[S]}),$$
$$\mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{B_1(o)\}\rangle}}(o, w_{[S']}) \subseteq \mathbf{r}_{\Sigma_2}^{\mathcal{G}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{B_1(o)\}\rangle}}(o, w_{[S]}).$$

It implies, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B_1 \sqsubseteq_{\rightsquigarrow} \exists S$. Further, since $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq B_1$, we have that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists Q$ for some role Q (possibly coinciding with S) such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models Q \sqsubseteq S$. It is straightforward to verify that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{Q \sqsubseteq Q',\ \exists Q^- \sqsubseteq C'\}$. $\square$

Proposition 7.2.14. *For a role R over Σ_1 and Q' over Σ_2, inclusion $R \sqsubseteq Q'$ is* representable *in $\mathcal{T}_1$ and $\mathcal{M}$ if and only if there exists R' over Σ_2 s.t. $\mathcal{T}_{12} \models R \sqsubseteq R'$, and*

(RINCL) *for each $\mathcal{T}_1$-consistent role S over Σ_1, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R'$ implies $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq Q'$;*

(EXINCL) *B', C' satisfy conditions (cincl) and (mrole) for $B' = \exists R'$, $C' = \exists Q'$, and $B' = \exists R'^-$, $C' = \exists Q'^-$.*

Then, $R \sqsubseteq Q'$ is representable by *$R' \sqsubseteq Q'$.*

Proof. ($\Leftarrow$) Let R be a role over Σ_1 and Q' over Σ_2, $R' \neq Q'$, and conditions (rincl) and (exincl) are satisfied. We show inclusion $R \sqsubseteq Q'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ by $R' \sqsubseteq Q'$. Similarly, to the proof of Proposition 7.2.13, take $\mathcal{T}_2$ a reserved UCQ-representation for $\mathcal{T}_1$ under $\mathcal{M}$: we prove $\mathcal{T}_2' = \mathcal{T}_2 \cup \{R' \sqsubseteq Q'\}$ is a reserved UCQ-representation by showing the direction of condition (i) stating that for each $\mathcal{T}_1$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent pair of concepts or roles (X, Y), (X, Y) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, the $\Leftarrow$ direction of condition (ii), and condition (iv) of Lemma 7.1.1 are satisfied.

Satisfaction of conditions (ii) and (i) of Lemma 7.1.1 can be shown by analogy with the corresponding proofs in Proposition 7.2.13. Note, here for concept inclusions/disjointness assertions we use the fact that $\exists R', \exists Q'$ and $\exists R'^-, \exists Q'^-$ satisfy (cincl), and for role inclusions/disjoint-ness assertions we use the fact R', Q' satisfy (rincl).

For condition (iv), the interesting case to consider is $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D \sqsubsetneqq \exists R'$, with D a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept over Σ_1, such that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}(w_{[R]}) = \{\exists R'^-\} \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_2 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}(o, w_{[R]}) = \{R'\}.$$

Now, given $R' \sqsubseteq Q' \in \mathcal{T}_2'$, we have that

$$\mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}(w_{[R]}) \supseteq \{\exists R'^-, \exists Q'^-\} \text{ and } \mathbf{r}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_2' \cup \mathcal{T}_{12}, \{D(o)\}\rangle}(o, w_{[R]}) \supseteq \{R', Q'\}.$$

By condition (ii), it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \exists R'$. As $\exists R'^-$ and $\exists Q'^-$ satisfy (mrole) and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \exists R'$, it follows there exists S such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubsetneqq \exists S$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S \sqsubseteq R', \exists S^- \sqsubseteq \exists Q'^-\}$; moreover by $\exists Q'^- \neq \exists R'^-$ and the structure of $\mathcal{T}_1 \cup \mathcal{T}_{12}$ it follows S is over Σ_1. From the latter we obtain a role Q over Σ_1 such that $\mathcal{T}_1 \models S \sqsubseteq Q$ and $\mathcal{T}_{12} \models Q \sqsubseteq R$, moreover $\exists Q^-$ and Q are $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Now, assume $\mathcal{T}_2 \models \exists R'^- \sqsubseteq E'$; then $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \exists Q^- \sqsubseteq E'$, and since $\mathcal{T}_2$ satisfies condition (ii) it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists Q^- \sqsubseteq E'$, therefore $E' \in \mathbf{t}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}(w_{[S]})$. Similarly, for T' such that $\mathcal{T}_2 \models R' \sqsubseteq T'$, we can show $T' \in \mathbf{r}_{\Sigma_2}^{\mathcal{G}\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \{D(o)\}\rangle}(o, w_{[S]})$. Thus, we take $y = w_{[S]}$ to satisfy condition (iv) of Lemma 7.1.1.

($\Rightarrow$) Suppose inclusion $R \sqsubseteq Q'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$ by a target axiom α. Then $\mathcal{T}_2 = \{\alpha\}$ is a reserved UCQ-representation and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models R \sqsubseteq Q'$. If $\mathcal{T}_{12} \models R \sqsubseteq Q'$, we take R' equal to Q': obviously, (rincl) and (exincl) are satisfied. Now, assume $\mathcal{T}_{12} \not\models R \sqsubseteq Q'$. Then it must be the case α is of the form $S' \sqsubseteq Q'$ and $\mathcal{T}_{12} \models R \sqsubseteq S'$ for some role S' over Σ_2. So we take R' equal to S', then (rincl) is shown similarly to (cincl) in the proof of Proposition 7.2.13, and satisfaction of (exincl) is shown exactly as in the proof of $\Rightarrow$ of Proposition 7.2.13 for $B' = \exists R'$, $C' = \exists Q'$, and $B' = \exists R'^-$, $C' = \exists Q'^-$. ☐

Proposition 7.2.15. *For roles R_1, R_2 over Σ_1, (R_1, R_2) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$ iff* ***either*** *for $\{R, Q\} \subseteq \{R_1, R_2\}$ there exists R' over Σ_2 such that*

(A) *$\mathcal{T}_{12} \models R \sqsubseteq R'$, and $Q \sqsubseteq \neg R' \in \mathcal{T}_{12}$, or there is Q' over Σ_2 s.t. $\mathcal{T}_{12} \models Q \sqsubseteq Q'$ and*

(RINCONS) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent pair of roles S_1, S_2 over Σ_1 it is* not the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S_1 \sqsubseteq R',\ S_2 \sqsubseteq Q'\}$;

(NOGEN) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept D over Σ_1 and each role S such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists S$, it is* neither the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R' \sqcap Q'$, nor $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R'^- \sqcap Q'^-$,

(B) *or $\mathcal{T}_{12} \models R \sqsubseteq \neg R'$ and inclusion $Q \sqsubseteq R'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$;*

or *for $\{B, C\} \subseteq \{\exists R_1, \exists R_2\}$ or $\{\exists R_1^-, \exists R_2^-\}$ there exists B' over Σ_2 such that*

(C) *$\mathcal{T}_{12} \models B \sqsubseteq B'$, and $C \sqsubseteq \neg B' \in \mathcal{T}_{12}$, or there is C' over Σ_2 s.t. $\mathcal{T}_{12} \models C \sqsubseteq C'$ and*

(CINCONS) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent pair of concepts D_1, D_2 over Σ_1 it is* not the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{D_1 \sqsubseteq B',\ D_2 \sqsubseteq C'\}$;

(NOREND) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept D over Σ_1 and each role S such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\rightsquigarrow} \exists S$ it is* not the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists S^- \sqsubseteq B' \sqcap C'$,

(D) *or $\mathcal{T}_{12} \models B \sqsubseteq \neg B'$ and inclusion $C \sqsubseteq B'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$;*

Then (R_1, R_2) is target contradictable by either $R' \sqsubseteq R'$, or $Q' \sqsubseteq \neg R'$ in (A), by axiom α, where $Q \sqsubseteq R'$ is representable by α in (B), by either $B' \sqsubseteq B'$, or $C' \sqsubseteq \neg B'$ in (C), and by axiom α, where $C \sqsubseteq B'$ is representable by α in (D).

Proof. ($\Leftarrow$) Let R_1, R_2 be roles over Σ_1 and one of the conditions (A), (B), (C), or (D) is satisfied. We show (R_1, R_2) is target contradictable by α given by each of the conditions. Take $\mathcal{T}_2$ a reserved UCQ-representation for $\mathcal{T}_1$ under $\mathcal{M}$: we prove $\mathcal{T}_2' = \mathcal{T}_2 \cup \{\alpha\}$ is a reserved UCQ-representation, by showing conditions (i), (ii), and (iv) of Lemma 7.1.1 are satisfied (only the required directions, see the proof of Proposition 7.2.13). That (R_1, R_2) is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent, follows immediately from the shape of α and $\mathcal{T}_{12}$ in each of the cases. Observe that if α is given by one of the conditions (B) or (D), then $\mathcal{T}_2'$ is a reserved UCQ-representation follows from the proof of Propositions 7.2.13 and 7.2.14. As for α given by conditions (A) or (C), it should be clear that conditions (ii) and (iv) of Lemma 7.1.1 are satisfied, as disjointness assertions do not affect entailments of the concept and role inclusions. Therefore, below we show $\mathcal{T}_2'$ satisfies condition (i).

Assume condition (A) is satisfied, and $\alpha = Q' \sqsubseteq \neg R'$ (the case $\alpha = R' \sqsubseteq R'$ is trivial), hence $\mathcal{T}_{12} \not\models Q \sqsubseteq \neg R'$. Let (D_1, D_2) be a pair of $\mathcal{T}_1$-consistent, $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent concepts. The case both D_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-consistent is not possible due to the shape of α. Then some D_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent, and by (tinconsc) it follows there exist $n \geq 1$ and distinct roles $S_1', \ldots, S_n'$ such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \{D_i \sqsubseteq_{\rightsquigarrow} \exists S_1',\ \exists S_j'^- \sqsubseteq_{\rightsquigarrow} \exists S_{j+1}'\}$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S_n' \sqsubseteq R' \sqcap Q'$ or $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S_n' \sqsubseteq R'^- \sqcap Q'^-$. In the following, we consider only $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S_n' \sqsubseteq R' \sqcap Q'$.

For the sake of contradiction, assume D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. If $n = 1$ and S_1' is a role over Σ_1 (i.e., $D_i = \exists S_1'$ and S_1' is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent), then we obtain contradiction with (rincons) rised from the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. If $n = 1$ and S_1' is a role over Σ_2, then since $\mathcal{T}_2$ is a reserved UCQ-representation and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D_i \sqsubseteq_{\rightsquigarrow} \exists S_1'$, by condition (iv), we obtain a role S_1 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D_i \sqsubseteq_{\rightsquigarrow} \exists S_1$, and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S_1 \sqsubseteq R' \sqcap Q'$: contradiction with (nogen).

For $n > 1$, inductively using condition (iv), we obtain roles $S_1, \ldots, S_{n-1}$ over Σ_1 and S_n s.t. $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{D_i \sqsubseteq_{\rightsquigarrow} \exists S_1,\ \exists S_j^- \sqsubseteq_{\rightsquigarrow} \exists S_{j+1}\}$, and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S_n \sqsubseteq R' \sqcap Q'$. Then (nogen) implies that $\exists S_{n-1}^-$ is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Finally, we conclude that D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, hence (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent.

Let (S_1, S_2) be a pair of $\mathcal{T}_1$-consistent, $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent roles. For the sake of contradiction, assume (S_1, S_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent (and each of S_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent).

Suppose both S_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-consistent. From the shape of α, without loss of generality, we may assume that $\mathcal{T}_2' \cup \mathcal{T}_{12} \models \{S_1 \sqsubseteq R',\ S_2 \sqsubseteq Q'\}$. From condition (ii), we obtain $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S_1 \sqsubseteq R',\ S_2 \sqsubseteq Q'\}$, which contradicts (rincons).

Suppose one of S_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent. Then by (tinconsr) either $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S_i \sqsubseteq R' \sqcap Q'$ or $\mathcal{T}_2 \cup \mathcal{T}_{12} \models S_i \sqsubseteq R'^- \sqcap Q'^-$, or D is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent for $D = \exists S_i$ or $D = \exists S_i^-$. In the latter case, we obtain contradiction as in the case (D_1, D_2) is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent. In the former case, from condition (ii), it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S_i \sqsubseteq R' \sqcap Q'$ or $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S_i \sqsubseteq R'^- \sqcap Q'^-$, which contradicts (rincons). Finally, we conclude (S_1, S_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent.

Assume condition (C) is satisfied, and $\alpha = C' \sqsubseteq \neg B'$ (the case $\alpha = B' \sqsubseteq B'$ is trivial), hence $\mathcal{T}_{12} \not\models C \sqsubseteq \neg B'$. Let (D_1, D_2) be a pair of $\mathcal{T}_1$-consistent, $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent concepts. For the sake of contradiction, assume (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent (and each of D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent).

Suppose both D_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-consistent. From the shape of α, without loss of generality, we may assume that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \{D_1 \sqsubseteq B',\ D_2 \sqsubseteq C'\}$. From condition (ii), it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{D_1 \sqsubseteq B',\ D_2 \sqsubseteq C'\}$: contradiction with (cincons).

Suppose one of D_i is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent. By (tinconsc), consider $\mathcal{T}_2 \cup \mathcal{T}_{12} \models D_i \sqsubseteq B' \sqcap C'$. From condition (ii), it follows $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D_i \sqsubseteq B' \sqcap C'$: contradiction with (cincons). Now, consider the case there exist $n \geq 1$ and distinct roles $S_1', \ldots, S_n'$ such that $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \{D_i \sqsubseteq_{\rightsquigarrow} \exists S_1',\ \exists S_j'^- \sqsubseteq_{\rightsquigarrow} \exists S_{j+1}'\}$ and $\mathcal{T}_2 \cup \mathcal{T}_{12} \models \exists S_n'^- \sqsubseteq B' \sqcap C'$. Inductively using condition (iv), we obtain roles $S_1, \ldots, S_{n-1}$ over Σ_1 and S_n s.t. $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{D_i \sqsubseteq_{\rightsquigarrow} \exists S_1,\ \exists S_j^- \sqsubseteq_{\rightsquigarrow} \exists S_{j+1}\}$, and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists S_n^- \sqsubseteq B' \sqcap C'$. Then (norend) implies that $\exists S_{n-1}^-$ (or D_i if $n = 1$) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts the assumption D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Finally, we conclude that D_i is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, hence (D_1, D_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent.

Let (S_1, S_2) be a pair of $\mathcal{T}_1$-consistent, $\mathcal{T}_2 \cup \mathcal{T}_{12}$-consistent and $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent roles. From the shape of α, it follows D is $\mathcal{T}_2' \cup \mathcal{T}_{12}$-inconsistent, for $D = \exists S_i$ or $D = \exists S_i^-$ and $i \in \{1, 2\}$. It can be shown D is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent as above.

($\Rightarrow$) Suppose pair (R_1, R_2) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$ by a target axiom α. If (R_1, R_2) is $\mathcal{T}_{12}$-inconsistent, then there exist $R, Q \in \{R_1, R_2\}$ and R' over Σ_2 such that $\mathcal{T}_{12} \models \{R \sqsubseteq R',\ Q \sqsubseteq \neg R'\}$ (hence, (A) is satisfied), or there exist B, C in $\{\exists R_1, \exists R_2\}$ or in $\{\exists R_1, \exists R_2\}$ and B' over Σ_2 such that $\mathcal{T}_{12} \models \{B \sqsubseteq B',\ C \sqsubseteq \neg B'\}$ (hence, (C) is satisfied).

Assume (R_1, R_2) is $\mathcal{T}_{12}$-consistent. Then α is a non-trivial axiom, $\mathcal{T}_2 = \{\alpha\}$ is a reserved UCQ-representation, and (R_1, R_2) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent.

Suppose α is a role disjointness assertion $S_1 \sqsubseteq \neg S_2$. Then it follows there exist $R, Q \in \{R_1, R_2\}$ and $S, T \in \{S_1, S_2\}$ such that $\mathcal{T}_{12} \models \{R \sqsubseteq S,\ Q \sqsubseteq T\}$. So we set R' equal to S and Q' equal to T. We prove (rincons) and (nogen) are satisfied. For (rincons), assume a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent pair of roles S_1, S_2 over Σ_1 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S_1 \sqsubseteq R',\ S_2 \sqsubseteq Q'\}$. It follows there exist S_{11}, S_{22} over Σ_1 such that $\mathcal{T}_1 \models \{S_1 \sqsubseteq S_{11},\ S_2 \sqsubseteq S_{22}\}$ and $\mathcal{T}_{12} \models \{S_{11} \sqsubseteq R',\ S_{22} \sqsubseteq Q'\}$. Next, (S_{11}, S_{22}) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, and since $\mathcal{T}_2$ is a reserved UCQ-representation, it follows (S_{11}, S_{22}) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts (S_1, S_2) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Hence, it cannot be the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{S_1 \sqsubseteq R',\ S_2 \sqsubseteq Q'\}$. For (nogen), assume a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept D over Σ_1 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\leadsto} \exists S$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R' \sqcap Q'$. If S is over Σ_1, then as above, we obtain a contradiction with D being $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. If S is over Σ_2, it follows $S = R' = Q'$, and there exists a concept D_1 over Σ_1 such that $\mathcal{T}_1 \models D \sqsubseteq D_1$ and $\mathcal{T}_{12} \models D_1 \sqsubseteq \exists S$. As above, (D_1, D_1) is $\mathcal{T}_2 \cup \mathcal{T}_{12}$-inconsistent, and since $\mathcal{T}_2$ is a reserved UCQ-representation, it follows (D_1, D_1) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent, which contradicts D is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent. Hence, it cannot be the case $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R' \sqcap Q'$. In a similar way we obtain a contradiction if assume $\mathcal{T}_1 \cup \mathcal{T}_{12} \models S \sqsubseteq R'^- \sqcap Q'^-$. Thus, (A) is satisfied.

Suppose α is a role inclusion assertion $S_1 \sqsubseteq S_2$. Then it follows there exist $R, Q \in \{R_1, R_2\}$ such that $\mathcal{T}_{12} \models \{R \sqsubseteq \neg S_2,\ Q \sqsubseteq S_1\}$. So we set R' equal to S_2, the proof $Q \sqsubseteq R'$ is representable by $S_1 \sqsubseteq R'$ is similar to the proof of $\Rightarrow$ of Proposition 7.2.14. Thus, (B) is satisfied.

Suppose α is a concept disjointness assertion $D_1 \sqsubseteq \neg D_2$. Then there exist B, C in $\{\exists R_1, \exists R_2\}$ or $\{\exists R_1^-, \exists R_2^-\}$ and $D, E \in \{D_1, D_2\}$ such that $\mathcal{T}_{12} \models \{B \sqsubseteq D,\ C \sqsubseteq E\}$. So we set B' equal to D and C' equal to E. We can prove (cincons) and (norend) are satisfied by analogy with the proof of (rincons) and (nogen). Thus, (C) is satisfied.

Suppose α is a concept inclusion assertion $D_1 \sqsubseteq D_2$. Then it follows there exist B, C in $\{\exists R_1, \exists R_2\}$ or $\{\exists R_1^-, \exists R_2^-\}$ such that $\mathcal{T}_{12} \models \{B \sqsubseteq \neg D_2,\ C \sqsubseteq D_1\}$. So we set B' equal to D_2, the proof $C \sqsubseteq B'$ is representable by $D_1 \sqsubseteq B'$ is similar to the proof of $\Rightarrow$ of Proposition 7.2.13. Thus, (D) is satisfied. □

Proposition 7.2.16. *For concepts B_1, B_2 over Σ_1, (B_1, B_2) is* target contradictable *in $\mathcal{T}_1$ and $\mathcal{M}$ if* ***either*** *for $\{B, C\} \subseteq \{B_1, B_2\}$ there exists B' over Σ_2 such that*

(E) *$\mathcal{T}_{12} \models B \sqsubseteq B'$, and either $C \sqsubseteq \neg B' \in \mathcal{T}_{12}$, or there exists C' over Σ_2 s.t. $\mathcal{T}_{12} \models C \sqsubseteq C'$ and*

- (CINCONS) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent pair of concepts D_1, D_2 over Σ_1 it is* not the case *$\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{D_1 \sqsubseteq B',\ D_2 \sqsubseteq C'\}$;*
- (NOREND) *for each $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept D over Σ_1 and each role S such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq_{\leadsto} \exists S$ it is* not the case *$\mathcal{T}_1 \cup \mathcal{T}_{12} \models \exists S^- \sqsubseteq B' \sqcap C'$,*

(F) *or $\mathcal{T}_{12} \models B \sqsubseteq \neg B'$ and inclusion $C \sqsubseteq B'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$;*

or *$B_1 = \exists R$ or $B_2 = \exists R$ for a role R, and*

(G) *(R, R) is* target contradictable *in $\mathcal{T}_1$ and $\mathcal{M}$.*

Then (B_1, B_2) is target contradictable by either $B' \sqsubseteq B'$ or $C' \sqsubseteq \neg B'$ in (E), by axiom α, where $C \sqsubseteq B'$ is representable by α in (F), and by axiom α such that (R, R) is target contradictable by α in (G).

Proof. The proof is similar to the proof of Proposition 7.2.15. □

This concludes the proof of Theorem 7.2.12. □

7.2.1 *Computing* UCQ-*representations*

We conclude with an algorithm for computing a UCQ-representation for a given source TBox $\mathcal{T}_1$ and a mapping $\mathcal{M}$, which can be extracted from the proof of Lemma 7.2.8. The algorithm is presented in Figure 14.

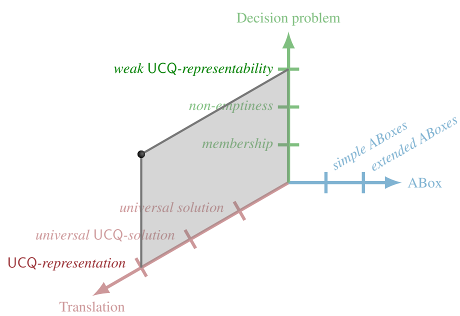
Algorithm ComputeUCQRepresentation$(\mathcal{M}, \mathcal{T}_1)$
Input: mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and TBox $\mathcal{T}_1$ over Σ_1
Output: TBox $\mathcal{T}_2$ over Σ_2 if $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}$, **nothing** otherwise.

$\mathcal{T}_2 := \{\}$.
for each $\mathcal{T}_1$-consistent pair of concepts B, C over Σ_1
  if (B, C) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent
    if (B, C) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$
      $\mathcal{T}_2 = \mathcal{T}_2 \cup \{\alpha\}$, where (B, C) is target contradictable by α.
    else return **nothing**.
for each $\mathcal{T}_1$-consistent pair of roles R, S over Σ_1
  if (R, S) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent
    if (R, S) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$
      $\mathcal{T}_2 := \mathcal{T}_2 \cup \{\alpha\}$, where (R, S) is target contradictable by α.
    else return **nothing**.
for each $\mathcal{T}_1$-consistent concept B over Σ_1 and concept B' over Σ_2
  if $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$
    if $B \sqsubseteq B'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$
      $\mathcal{T}_2 := \mathcal{T}_2 \cup \{\alpha\}$, where $B \sqsubseteq B'$ is representable by α.
    else return **nothing**.
for each $\mathcal{T}_1$-consistent role R over Σ_1 and role R' over Σ_2
  if $\mathcal{T}_1 \cup \mathcal{T}_{12} \models R \sqsubseteq R'$
    if $R \sqsubseteq R'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}$
      $\mathcal{T}_2 := \mathcal{T}_2 \cup \{\alpha\}$, where $R \sqsubseteq R'$ is representable by α.
    else return **nothing**.
return $\mathcal{T}_2$.

Figure 14: Algorithm ComputeUCQRepresentation.

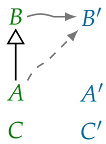
7.3 WEAK UCQ-REPRESENTABILITY

In this final section we show how the weak UCQ-representability problem can be solved in NLogSpace-complete relying on the procedure for checking the non-emptiness problem.



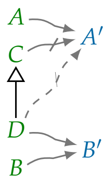
We start with some example of weakly UCQ-representable TBoxes.

Example 7.3.1. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), B(\cdot), C(\cdot)\}$, $\Sigma_2 = \{A'(\cdot), B'(\cdot), C'(\cdot)\}$ and $\mathcal{T}_{12} = \{B \sqsubseteq B'\}$. Moreover, let $\mathcal{T}_1 = \{A \sqsubseteq B\}$. It was shown in Example 4.3.1-(1), that $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$. By the characterization in Lemma 7.2.8, condition (II) is violated: there is no way to imply by a target TBox and $\mathcal{T}_{12}$ the axiom $A \sqsubseteq B'$ implied by $\mathcal{T}_1 \cup \mathcal{T}_{12}$. So one can simply enrich the mapping by exactly this inclusion: consider $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$, where $\mathcal{T}_{12}^\star = \mathcal{T}_{12} \cup \{A \sqsubseteq B'\}$, then $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}^\star$ and $\mathcal{T}_2 = \{\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}^\star$.



Hence $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}$. ■

Example 7.3.2. Assume that $\mathcal{M} = (\{A(\cdot), B(\cdot), C(\cdot), D(\cdot)\}, \{A'(\cdot), B'(\cdot)\}, \mathcal{T}_{12})$, where $\mathcal{T}_{12} = \{A \sqsubseteq A', B \sqsubseteq B', C \sqsubseteq \neg A', D \sqsubseteq B'\}$, and let $\mathcal{T}_1 = \{D \sqsubseteq C\}$. We explained in Example 7.1.6 why $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$: condition (I) of Lemma 7.2.8 is violated for the pair (A, D), which is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. Note, that there is no problem for the pair (A, C), which is actually $\mathcal{T}_{12}$-inconsistent. From (A, D) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent it follows that for some concept D' over Σ_2, $\mathcal{T}_1 \cup \mathcal{T}_{12} \models A \sqsubseteq D'$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models D \sqsubseteq \neg D'$: in this case $D' = A'$ and $\mathcal{T}_{12} \models A \sqsubseteq A'$. So we can add the axiom $D \sqsubseteq \neg A'$ to the mapping to achieve that (A, D) is inconsistent with respect to the mapping alone. Thus, consider $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$, where $\mathcal{T}_{12}^\star = \mathcal{T}_{12} \cup \{D \sqsubseteq \neg A'\}$, then $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}^\star$ and $\mathcal{T}_2 = \{\}$ is a UCQ-representation of $\mathcal{T}_1$ under $\mathcal{M}^\star$.



Hence $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}$. ■

The following example illustrates a case of a TBox that is not weakly UCQ-representable.

Example 7.3.3. Let $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot)\}, \Sigma_2 = \{A'(\cdot), B'(\cdot)\}$, and $\mathcal{T}_{12} = \{A \sqsubseteq A', \exists R^- \sqsubseteq B'\}$. Furthermore, assume that $\mathcal{T}_1 = \{A \sqsubseteq \exists R\}$. In Example 4.3.2 we showed that $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}$, now we will show that is not weakly UCQ-representable either. Recall that in the definition of weak UCQ-representability, we are looking for an extended mapping $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ such that $\mathcal{T}_{12} \subseteq \mathcal{T}_{12}^\star$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^\star$, and in this example, the maximal such $\mathcal{T}_{12}^\star$ coincides with $\mathcal{T}_{12}$ itself. Therefore, there exists no $\mathcal{M}^\star$ as above such that $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}^\star$, and $\mathcal{T}_{12}$ is not weakly UCQ-representable under $\mathcal{M}$.

These three examples give an idea of how it is possible to check weak UCQ-representability.

Lemma 7.3.4. *Given a mapping $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ and a TBox $\mathcal{T}_1$ over Σ_1, such that $\mathcal{T}_1$ is not* UCQ*-representable under $\mathcal{M}$. Let $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$ be a mapping such that $\mathcal{T}_{12}^\star$ is the maximal (with respect to set-inclusion) TBox from Σ_1 to Σ_2 satisfying $\mathcal{T}_{12} \subseteq \mathcal{T}_{12}^\star$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^\star$. Then $\mathcal{T}_1$ is weakly* UCQ*-representable under $\mathcal{M}$ if and only if $\mathcal{T}_1$ and $\mathcal{M}^\star$ satisfy condition* (III) *of Lemma* 7.2.8*.*

Proof. ($\Leftarrow$) Assume $\mathcal{T}_1$ and $\mathcal{M}^\star$ satisfy condition (III). We show conditions (I) and (II) are satisfied for $\mathcal{T}_1$ and $\mathcal{M}^\star$.

(I) Assume (B, C) is NOT target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$, for a $\mathcal{T}_1$-consistent pair of concepts B, C over Σ_1, such that (B, C) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. Let D' be a concept over Σ_2 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{B \sqsubseteq D', C \sqsubseteq \neg D'\}$ or $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{B \sqsubseteq \neg D', C \sqsubseteq D'\}$. Such D' exists because (B, C) is $\mathcal{T}_1$-consistent and $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. Then by construction of $\mathcal{T}_{12}^\star$, it contains inclusions $B \sqsubseteq D'$, $C \sqsubseteq \neg D'$ in the former case, and $B \sqsubseteq \neg D'$, $C \sqsubseteq D'$ in the latter case. So, (B, C) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}^\star$.

(II) Assume (R, Q) is NOT target contradictable in $\mathcal{T}_1$ and $\mathcal{M}$, for a $\mathcal{T}_1$-consistent pair of roles R, Q over Σ_1, such that (R, Q) is $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. Let S' be a role over Σ_2 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{R \sqsubseteq S', Q \sqsubseteq \neg S'\}$ or $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \{R \sqsubseteq \neg S', Q \sqsubseteq S'\}$. Such S' exists because (R, Q) is $\mathcal{T}_1$-consistent and $\mathcal{T}_1 \cup \mathcal{T}_{12}$-inconsistent. Then $\mathcal{T}_{12}^\star$ contains inclusions $R \sqsubseteq S'$, $Q \sqsubseteq \neg S'$ in the former case, and $R \sqsubseteq \neg S'$, $Q \sqsubseteq S'$ in the latter case. So, (R, Q) is target contradictable in $\mathcal{T}_1$ and $\mathcal{M}^\star$.

(III) Assume inclusion $B \sqsubseteq B'$ is NOT representable in $\mathcal{T}_1$ and $\mathcal{M}$, for a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent concept B over Σ_1 and a concept B' over Σ_2 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models B \sqsubseteq B'$. Then $\mathcal{T}_{12}^\star$ contains $B \sqsubseteq B'$, so $B \sqsubseteq B'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}^\star$.

(IV) Assume inclusion $R \sqsubseteq R'$ is NOT representable in $\mathcal{T}_1$ and $\mathcal{M}$, for a $\mathcal{T}_1 \cup \mathcal{T}_{12}$-consistent role R over Σ_1 and a role R' over Σ_2 such that $\mathcal{T}_1 \cup \mathcal{T}_{12} \models R \sqsubseteq R'$. Then $\mathcal{T}_{12}^\star$ contains $R \sqsubseteq R'$, so $R \sqsubseteq R'$ is representable in $\mathcal{T}_1$ and $\mathcal{M}^\star$.

Therefore, $\mathcal{T}_1$ is UCQ-representable under $\mathcal{M}^\star$, hence $\mathcal{T}_1$ is weakly UCQ-representable under $\mathcal{M}$.

($\Rightarrow$) Assume condition (III) of Lemma 7.2.8 is violated by $\mathcal{T}_1$ and $\mathcal{M}^\star$. Then by Lemma 7.2.8, $\mathcal{T}_1$ is not UCQ-representable under $\mathcal{M}^\star$. Since $\mathcal{M}^\star = (\Sigma_1, \Sigma_2, \mathcal{T}_{12}^\star)$, and $\mathcal{T}_{12}^\star$ is the maximal TBox that satisfies $\mathcal{T}_{12} \subseteq \mathcal{T}_{12}^\star$ and $\mathcal{T}_1 \cup \mathcal{T}_{12} \models \mathcal{T}_{12}^\star$, we conclude $\mathcal{T}_1$ is not weakly UCQ-representable under $\mathcal{M}$. □

Therefore, we obtain the following complexity bound.

Theorem 7.3.5. *The weak* UCQ*-representability problem is* NLOGSPACE*-complete.*

We conclude with the case of *DL-Lite*$_{RDFS}$. Interestingly, in this case for a source TBox $\mathcal{T}_1$ and a mapping $\mathcal{M}$, condition (III) of Lemma 7.2.8 is never triggered as in *DL-Lite*$_{RDFS}$ it is not possible to generate new objects. Therefore, this condition is trivially satisfied, so in *DL-Lite*$_{RDFS}$, it is always possible to represent the source implicit knowledge by enriching mappings.

Theorem 7.3.6. *In DL-Lite*$_{RDFS}$*, the weak* UCQ*-representability problem is in* TRIVIAL*.*

8 RELATED WORK

8.1 DATA EXCHANGE

Data exchange is the starting point for the knowledge base exchange problem, and the former is related to the latter mainly in two respects. First, the motivation for KB exchange originates from data exchange, and in particular from data exchange with incomplete information. Second, the definition of the KB exchange framework inherits many notions and definitions from the data exchange framework. Below we give a short introduction to data exchange and data exchange with incomplete information.

8.1.1 *Data Exchange with Complete Data*

Data exchange deals with transferring data between differently structured databases. A *schema* is a finite set $\mathbf{R} = \{R_1, \ldots, R_n\}$ of relation symbols R_i each with the associated arity n_i. Given a schema $\mathcal{R}$, an *instance* I of $\mathbf{R}$, assigns to each relation symbol R_i of $\mathbf{R}$ a finite n_i-ary relation R_i^I. The *domain* of an instance I, denoted by $dom(I)$, is the set of all elements that occur in any of the relations R_i^I, where each element can be either a *constant*, or a *labeled null*.

Let $\mathbf{S} = \{S_1, \ldots, S_n\}$ and $\mathbf{T} = \{T_1, \ldots, T_m\}$ be two disjoint schemas. Then a *schema mapping* is a triple $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ such that Σ is a finite set of constraints, (closed) formulas in a logic $\mathcal{L}$ over $\mathbf{S} \cup \mathbf{T}$. $\mathbf{S}$ is called the *source* schema, and $\mathbf{T}$ is called the *target* schema. Given a instance I of $\mathbf{S}$ (called source instance), the instance J of $\mathbf{T}$ (called target instance) is said to be a *solution* for I under $\mathcal{M}$, if (I, J) satisfies every sentence in Σ. The set of all solutions for I under $\mathcal{M}$ is denoted by $Sol_{\mathcal{M}}(I)$. The *data exchange problem* is given a mapping $\mathcal{M}$ and a source instance I, to find a solution for $\mathcal{I}$ under $\mathcal{M}$.

The class of constraints usually employed for data exchange are *tuple-generating dependencies (tgds)* and *equality-generating dependencies (egds)* that are first-order formulas of the form $\forall x \forall y(\phi(x, y) \to \exists z.\psi(x, z))$ and $\phi(x) \to (x_i = x_j)$ respectively, where $\phi(x, y)$ and $\psi(x, z)$ are conjunctions of atomic formulas. Moreover, for a source schema $\mathbf{S}$ and a target schema $\mathbf{T}$, a tgd $\forall x \forall y(\phi(x, y) \to \exists z.\psi(x, z))$ is called a *source-to-target tgd (st-tgd)* if ϕ is a formula over $\mathbf{S}$ and ψ is a formula over $\mathbf{T}$. Observe that, *DL-Lite* constraints are a subclass of so called *linear* tgds, which are in turn a subclass of *guarded* tgds (see, e.g., [66]).

Due to the existentially quantified variables on the right-hand side of tgds, it is possible to have (infinitely) many solutions. A natural question to ask then is which of them should be materialized. To this purpose, the concept of *universal solution* was introduced in [50] and it was argued that universal solutions are the preferred solutions in data exchange as they are the "most general" solutions: a solution J for I under $\mathcal{M}$ is said to be a *universal solution* for I under $\mathcal{M}$ if for every solution J' for I under $\mathcal{M}$, there exists a homomorphism $h : J \to J'$. Next, the "best" universal solution was identified in [52] as the smallest universal solution, which is the *core* of all universal solutions.

Most of results in data exchange focus on schema mappings of the form

$$(\mathbf{S}, \mathbf{T}, \Sigma_{st} \cup \Sigma_t),$$

where Σ_{st} is a set of st-tgds and Σ_t is a set of target tgds and egds. In addition, in many cases one considers a syntactic restriction on tgds: full and weakly acyclic tgds. Full tgds do not contain existentially quantified variables, thus exclude incompleteness in the target. The latter notion of weak acyclicity is widely adopted in data exchange as it assures termination of the chase, used to compute solutions. Below we present some of the known results on data exchange:

- There exists a data exchange setting $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st} \cup \Sigma_t)$, such that for a given source instance I, the problem of deciding whether I has a solution under $\mathcal{M}$, is undecidable.
- Let $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st} \cup \Sigma_t)$ be a fixed data exchange setting, such that Σ_t is the union of a set of target egds and a weakly acyclic set of target tgds.
  - There is a polynomial time algorithm such that for every source instance I, it first decides whether a solution for I exists, and if that is the case, it computes a universal solution for I in polynomial time using the chase [50];
  - There is a polynomial time algorithm that for every source instance I, computes the core of the universal solutions for I, if it exists [61];
  - Let q be a UCQ. Then the problem of computing certain answers for q under $\mathcal{M}$ can be solved in polynomial time [50];

8.1.2 *Data Exchange with Incomplete Data*

Recently there has been an interest in data exchange with data in the source incompletely specified [6, 12, 13]. A general framework for data exchange with incomplete data was proposed in [6]: in this setting the source data may be incompletely specified, and thus may represent (possibly infinitely) many source instances. This framework in based on the general notion of *representation system*: a *representation system* is a tuple $(\mathbf{R}, \textsc{Rep})$, where $\mathbf{R}$ is a set of *representatives* and $\textsc{Rep}$ is a function that assigns a set of instances to every element in $\mathbf{R}$. Intuitively, in terms of first-order logic, if $I \in \mathbf{R}$, then $\textsc{Rep}(I)$ is the set of all possible models of I.

Let $\mathcal{R} = (\mathbf{R}, \textsc{Rep})$ be a representation system, $\mathcal{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$ a mapping from a schema $\mathbf{S}$ to a schema $\mathbf{T}$, and I, J $\mathbf{R}$-elements of $\mathbf{S}$ and $\mathbf{T}$, respectively. For X a set of instances of $\mathbf{S}$, define $Sol_{\mathcal{M}}(X) = \bigcup_{I \in X} Sol_{\mathcal{M}}(I)$. Then J is said to be an *$\mathcal{R}$-solution* for I under $\mathcal{M}$ if $\textsc{Rep}(J) \subseteq Sol_{\mathcal{M}}(\textsc{Rep}(I))$, and J is said to be a *universal $\mathcal{R}$-solution* for I under $\mathcal{M}$ if $\textsc{Rep}(J) = Sol_{\mathcal{M}}(\textsc{Rep}(I))$.

The above mentioned definitions of solutions are extended in a natural way to the representation system of knowledge bases, where a knowledge base over a schema $\mathbf{S}$ is a pair (I, T) with I an instance of $\mathbf{S}$ and T a set of logical sentences over $\mathbf{S}$, and $\textsc{Rep}$ corresponds to $\textsc{Mod}$, the set of all possible models. We cite two results on knowledge exchange from [6], where tgd knowledge bases are being considered, that is, a tgd knowledge base is a pair (I, T), where I is an instance and T is a set of tgds.

- There exists a mapping $\mathcal{M} = (\mathbf{S}_1, \mathbf{S}_2, \Sigma_{12})$, with Σ_{12} a set of full st-tgds, for which the problem of verifying, given a tgd KB $\mathcal{K}_1$ over $\mathbf{S}_1$ and a tgd KB $\mathcal{K}_2$ over $\mathbf{S}_2$, whether $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$, is undecidable.
- Let $\mathcal{M} = (\mathbf{S}_1, \mathbf{S}_2, \Sigma_{12})$, with Σ_{12} a set of st-tgds. Then the problem of verifying, given a full-tgd KB $\mathcal{K}_1$ over $\mathbf{S}_1$ and a full-tgd KB $\mathcal{K}_2$ over $\mathbf{S}_2$, whether $\mathcal{K}_2$ is a solution for $\mathcal{K}_1$ under $\mathcal{M}$, is in PTIME.

Although, the KB exchange framework is largely based on the data exchange framework, it was not possible to adopt the techniques developed for the latter. There are two main reasons for that:

1) in KB exchange, we have to deal with the fact that a source KB represents (possibly infinitely) many actual models of possibly unbounded size, which then need to be "translated" according to the mapping. While in data exchange, there is one complete source instance for which it is straightforward to compute the translation (using the chase procedure).

2) Nevertheless, in the traditional data exchange setting, non-terminating chase (that generates an instance of unbounded size) can occur when trying to obtain a target instance satisfying the target constraints. Since in general the problem of existence of a solution is undecidable, to regain decidability one resorts to considering (weakly) acyclic sets of tgds, for which the chase procedure always terminates. In the context of *DL-Lite*$_{\mathcal{R}}$, which is already a very light-weight DL, one usually does not limit consideration to weakly acyclic TBoxes, as it is possible to obtain algorithms for the general case.

8.2 ONTOLOGY MODULARITY AND CONSERVATIVE EXTENSIONS

Modularity is another approach for collaborative ontology engineering and reuse of existing ontologies [39, 38, 101, 76]. The main idea of modularity is given an ontology $\mathcal{O}$ to split it into preferably small sub-ontologies, each of which can be used "autonomously" and independently of the rest of the ontology. Such sub-ontologies are called *modules*, and since they are typically of a small size (whereas the entire ontology could be huge), it is easier to understand them and to perform reasoning with them.

A logical framework for ontology modularity has been defined in [38] and is based on the notion of conservative extensions that provide the necessary formal means for defining and checking correctness of modules. Conservative extensions is a well-known notion in mathematical logic, where a logical theory T_2 is said to be a (proof theoretic) *conservative extension* of a theory T_1 if the language of T_2 extends the language of T_1; every theorem of T_1 is a theorem of T_2; and every theorem of T_2 that is in the language of T_1 is already a theorem of T_1.

In the Description Logics domain, conservative extensions have been shown to be necessary in the context of ontology modularity and ontology refinement [59, 36], so it led to rise of several works on conservative extensions in expressive DLs [86], in $\mathcal{EL}$ [85], and in *DL-Lite* [78, 79, 77].

While the data exchange setting influenced the decisions taken to define the KB exchange framework, conservative extensions are related to the KB exchange problem from the technical point of view:

- given two KBs $\mathcal{K}_1$ and $\mathcal{K}_2$, and a signature Σ, $\mathcal{K}_1$ Σ-*model entails* $\mathcal{K}_2$ if for each model $\mathcal{I}_1$ of $\mathcal{K}_1$ there exists a model $\mathcal{I}_2$ of $\mathcal{K}_2$ such that $\mathcal{I}_1$ and $\mathcal{I}_2$ agree on interpreted constants and on symbols from Σ, and $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-*model inseparable* if they Σ-model entail each other.

  One can see that $\mathcal{K}_2$ is a universal solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ iff $\mathcal{K}_2$ is Σ_2-model inseparable with $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$.

- given two KBs $\mathcal{K}_1$ and $\mathcal{K}_2$, and a signature Σ, $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$ if for each query q over Σ, $cert(q, \mathcal{K}_2) \subseteq cert(q, \mathcal{K}_1)$, and $\mathcal{K}_1$ and $\mathcal{K}_2$ are Σ-*query inseparable* if they Σ-query entail each other.

  Obviously, $\mathcal{K}_2$ is a universal UCQ-solution for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$ iff $\mathcal{K}_2$ is Σ_2-query inseparable with $\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle$.

- given two TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, and a signature Σ, $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ if for each ABox $\mathcal{A}$, $\langle \mathcal{T}_1, \mathcal{A} \rangle$ Σ-query entails $\langle \mathcal{T}_2, \mathcal{A} \rangle$, and $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-*query inseparable* if they Σ-query entail each other.

  Then, $\mathcal{T}_2$ and $\mathcal{T}_1 \cup \mathcal{T}_{12}$ are Σ_2-query inseparable implies $\mathcal{T}_2$ is a UCQ-representation for $\mathcal{K}_1 = \langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$.

  Note that TBox inseparability is a more general problem than UCQ-representability as in the former problem the quantification is over arbitrary ABoxes $\mathcal{A}$ (including those over Σ_2), while in the latter problem the quantification is over source ABoxes only (excluding those over Σ_2). Moreover, due to the particular shape of the TBox $\mathcal{T}_1 \cup \mathcal{T}_{12}$, there is a contrasting difference in the computational complexities of the two problems: UCQ-representability is NLOGSPACE-complete, while TBox CQ-inseparability is EXPTIME-complete.

Each of these problems reduces to checking some form of Σ-homomorphisms (finite or general). The techniques developed for inseparabilities can be used for solving the membership problems in the KB exchange framework and the other way around. In fact, we have shown in [23] the close connection between the membership problem for universal UCQ-solutions and the Σ-query inseparability problem (this result is out of the scope of this thesis). Currently, there are no known results for Σ-model entailment/inseparability in *DL-Lite*$_\mathcal{R}$, however we conjecture that the devised automata techniques can be used/adapted to this purpose.

8.3 ONTOLOGY ALIGNMENT

There exists a whole body of work on ontology alignment, also referred to as ontology mapping and ontology matching [34, 47, 99, 48]. The task of ontology alignment is given two ontologies to find correspondences between semantically related entities of ontologies resulting in a mapping between the ontologies. Such a problem arises in the contest of ontology merging, integration and alignment, which can be considered as an ontology reuse process.

There have been implemented several prototypes that allow for detecting and constructing mappings between ontologies such as PROMPT [94], SAMBO [82], Falcon [70], LogMap [71] and others. Moreover, in this community an ontology alignment evaluating initiative (OAEI) has been organized, where once a year the existing systems are being compared and evaluated [3, 99, 60].

Despite seeming similarity of the ontology alignment and knowledge base exchange problems, the former problem can be seen as the opposite of the latter, since in the KB exchange framework, a source KB and a mapping are assumed to be the input, and the task is to materialize a target KB, while in ontology alignment the input is two ontologies and the task is to find a mapping between them (which might not have a direction).

9

DISCUSSION

9.1 CONCLUSIONS

In this thesis we addressed the knowledge base exchange problem for Description Logic (DL) knowledge bases (KBs).

In Chapter 3 we defined the knowledge base exchange framework, where we specified three types of translations we would like to materialize: universal solutions, universal UCQ-solutions and UCQ-representations. Moreover, we defined the reasoning problems that we investigated in the rest of the thesis: the membership and the non-emptiness problems for each kind of translation.

Then, in Chapter 4 we discussed the basic properties of the three notions of solutions, and compared these to each other. There it became clear why we also need to consider extended ABoxes in the target.

In Chapter 5, we studied the complexity of KB exchange for universal solutions, and obtained various computational bounds for $DL\text{-}Lite_{\mathcal{R}}$ depending on the decision problem (membership or non-emptiness) and on the shape of target ABoxes (simple or extended):

- When only simple ABoxes are allowed in the target, we showed that both the membership and non-emptiness problems for universal solutions are solvable in polynomial time by providing a reduction to reachability games on graphs, and proved that this bound is tight.
- In the case where extended ABoxes are allowed in the target, we proved that the membership problem becomes NP-complete and identified labeled nulls as the source of complexity. In fact, even if the source TBox is empty, the problem remains NP-hard. Moreover, we provided a polynomial space lower bound and an exponential time algorithm based on two-way alternating automata for the non-emptiness problem in the latter case.

Such a discrepancy of results although surprising at first sight, can be easily justified. First of all, restricting attention to simple ABoxes reduces the space of all target ABoxes that need to be considered (actually, a finite number of them). Second, checking existence of a homomorphism from a tree to a graph formed by constants, which is required for solving the membership problem, is easy. On the other hand, the number of all possible extended target ABoxes is infinite, and checking existence of a homomorphism from a(n infinite) tree to a forest generated by constants is hard. As for $DL\text{-}Lite_{RDFS}$ KBs and mappings, we showed that the non-emptiness problem is trivially always true, and the membership problem is NLogSpace and NP-complete for simple and extended ABoxes respectively.

Next, in Chapter 6, we studied the complexity of KB exchange for universal UCQ-solutions, and showed that both the membership problem for simple target ABoxes and

Universal solutions	simple ABoxes	extended ABoxes
Membership	PTIME-complete	NP-complete
Non-emptiness	PTIME-complete	PSPACE-hard, in EXPTIME

Universal UCQ-solutions	simple ABoxes	extended ABoxes
Membership	PSPACE-hard	in EXPTIME
Non-emptiness	in EXPTIME	PSPACE-hard

UCQ-representations	Complexity
Membership	NLOGSPACE-complete
Non-emptiness	NLOGSPACE-complete
Weak UCQ-representability	NLOGSPACE-complete

Table 3: Complexity results for the membership and non-emptiness problems in $\textit{DL-Lite}_{\mathcal{R}}$.

the non-emptiness problem for extended target ABoxes are PSPACE-hard. As for $\textit{DL-Lite}_{RDFS}$ KBs and mappings, it turned out that the complexity of universal UCQ-solutions coincides with the complexity of universal solutions.

Finally, in Chapter 7, we studied the complexity of computing UCQ-representations, and obtained that UCQ-representations are the translations that are the simplest from the computational point of view.

- we developed graph-theoretic techniques for checking the membership and non-emptiness problems in NLOGSPACE;
- we showed that these two problems are NLOGSPACE-hard;
- we also obtained that the weak UCQ-representability problem is NLOGSPACE-complete.

These bounds hold for $\textit{DL-Lite}_{\mathcal{R}}$, and in the case of $\textit{DL-Lite}_{RDFS}$ there is only one difference: weak UCQ-representability is trivially true for any input TBox and mapping.

The summary of the results is once again presented in Tables 3 and 4.

9.2 WHAT IS A PREFERRED SOLUTION?

Out of the three notions of translations defined in Chapter 3, there is no single translation that could be unequivocally "preferred" in all possible scenarios. Each notion has its strengths and its weaknesses, which can be summarized as in Table 5.

Unquestionably, if one is interested in preserving logical correctness of the knowledge stored in the target KB, then universal solutions is the preferred translation: universal solutions are the most precise, model-theoretical translations. However, they present several limitations from the practical point of view:

- If one considers extended ABoxes, then universal solutions can be of exponential size.

Universal solutions	simple ABoxes	extended ABoxes
Membership	NLOGSPACE-complete	NP-complete
Non-emptiness	TRIVIAL	TRIVIAL

Universal UCQ-solutions	simple ABoxes	extended ABoxes
Membership	NLOGSPACE-complete	NP-complete
Non-emptiness	TRIVIAL	TRIVIAL

UCQ-representations	Complexity
Membership	NLOGSPACE-complete
Non-emptiness	NLOGSPACE-complete
Weak UCQ-representability	TRIVIAL

Table 4: Complexity results for the membership and non-emptiness problems in *DL-Lite*$_{RDFS}$.

- Universal solutions are sensitive to presence of disjointness assertions: in some cases one disjointness assertion is enough to ruin existence of a universal solution (see Example 4.1.7). Moreover, as will be discussed in Section 9.3, disjointness assertions complicate the characterization of universal solutions when more expressive target ABoxes are considered.
- Universal solutions are sensitive to whether the UNA is employed or not: there are examples when a universal solution exists under the UNA, but does not exist without the UNA.

Instead, if one considers a scenario where the main/only reasoning task is query answering over the target KB, then the query-based notions of translations become of particular interest. From Table 5 one can see that universal UCQ-solutions are more robust and behave better than universal solutions: they are usually more compact (in the worst case of the same size), and exist more often than universal solutions. Moreover, universal UCQ-solutions, in contrast with universal solutions and UCQ-representations, exist in every simple case, that is, a universal UCQ-solution exists whenever $\mathcal{T}_1 \cup \mathcal{T}_{12}$ is sufficiently simple (when $\mathcal{A}_1$ is given, it means that $\mathcal{U}_{\langle \mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1 \rangle}$ is finite). All this makes universal UCQ-solutions a good candidate for the preferred translation.

Lastly, in a scenario where data is changing or is not known, and the main reasoning task is query answering, UCQ-representations immediately appeal with their nice computational properties: they are decidable in polynomial time and their size is bound by a polynomial. When a UCQ-representation exists, it provides a straightforward poly-

	logically implied	compact	computationally simple	robust to UNA	robust to disjointness	exist in simple cases
universal solutions	YES	NO	YES/NO	NO	NO	NO/YES
universal UCQ-solutions	NO	YES/NO	NO	YES	YES	YES
UCQ-representations	NO	YES	YES	YES	YES	NO

Table 5: Properties of the three types of translation.

nomial time algorithm for computing universal UCQ-solutions of polynomial size. The main obstacle, however, for defining UCQ-representations to be the preferred translation is their "fragility": UCQ-representations do not exist even in some very minimalistic settings (see Example 4.3.1). Fortunately, this problem can be "fixed" by allowing for weak UCQ-representations and employing a reasonable assumption that usually roles get translated by the mapping (i.e., for "almost" each role R_1 over Σ_1 there exists an assertion of the form $R_1 \sqsubseteq R_2$ in the mapping).

Putting everything together, we conclude that the query-based notions of translations (universal UCQ-solutions and UCQ-representations) are probably the more promising translations to be materialized and used in practice.

9.3 OPEN PROBLEMS

We conclude with some open problems and future work.

To begin with, from the summary of the results one can see that some bounds are not tight or are missing. We note that it is already known that the membership problem for universal UCQ-solutions with simple target ABoxes is ExpTime-complete, but the lower bound was not included in the thesis as it was not obtained in of the scope of the thesis. As for the non-emptiness problem, we explained how to obtain a naive ExpTime algorithm, however we do not have any lower bound. In fact, this problem could turn out to be much simpler than ExpTime. It still remains to investigate the complexity of universal UCQ-solutions when extended ABoxes are allowed in the target.

Then, a straightforward extension to our work is to consider additional types of extended ABoxes. For instance, it is easy to see that allowing for inequalities between terms (e.g., $a \neq b$ in Example 4.1.7) and for negated atoms in the (target) ABox would allow one to obtain more universal solutions. However doing so would require employing some new techniques, as it immediately leads to some counter-intuitive examples:

Example 9.3.1. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where $\Sigma_1 = \{A(\cdot), R(\cdot,\cdot)\}$, $\Sigma_2 = \{S(\cdot,\cdot)\}$ and $\mathcal{T}_{12} = \{R \sqsubseteq \neg S,\ A \sqsubseteq \neg\exists S\}$. Moreover, let $\mathcal{T}_1 = \{A \sqsubseteq \exists R, \exists R^- \sqsubseteq \exists R\}$ and $\mathcal{A}_1 = \{A(a)\}$. Then $\mathcal{A}_2 = \{\neg\exists S(a), \neg S(a,a)\}$ is a universal solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$.

Now, if we imagine that the canonical model can have negated labels and the definition of homomorphism is extended accordingly, this example is counter-intuitive as $\mathcal{A}_2$ does not satisfy the characterization of universal solutions in Lemma 5.1.2. ∎

In the following example, we have an inequality between constants in the source ABox.

Example 9.3.2. Assume that $\mathcal{M} = (\Sigma_1, \Sigma_2, \mathcal{T}_{12})$, where

$$\begin{aligned}
\Sigma_1 &= \{A(\cdot), R(\cdot,\cdot), S(\cdot,\cdot), B(\cdot), C(\cdot)\},\\
\Sigma_2 &= \{B'(\cdot), C'(\cdot)\},\\
\mathcal{T}_{12} &= \{\exists R^- \sqsubseteq B',\ \exists S^- \sqsubseteq C',\ B \sqsubseteq B',\ C \sqsubseteq C'\}.
\end{aligned}$$

Moreover, let

$$\begin{aligned}
\mathcal{T}_1 &= \{A \sqsubseteq \exists R,\ A \sqsubseteq \exists S,\ \exists R^- \sqsubseteq \neg\exists S^-\}, \text{ and}\\
\mathcal{A}_1 &= \{A(a),\ B(b),\ C(c),\ C(d),\ c \neq d\}.
\end{aligned}$$

Then $\mathcal{A}_2 = \{B'(b), C'(c), C'(d), c \neq d\}$ is a universal solution for $\langle \mathcal{T}_1, \mathcal{A}_1 \rangle$ under $\mathcal{M}$. In fact, it is easy to see that for each model $\mathcal{I}$ of $\mathcal{K}_1$ and each $\mathcal{J}$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$, $\mathcal{J}$ is a model of $\mathcal{A}_2$. And on the other hand, for every model $\mathcal{J}$ of $\mathcal{A}_2$, there exists a model $\mathcal{I}$ of $\mathcal{K}_1$ such that $(\mathcal{I}, \mathcal{J}) \models \mathcal{T}_{12}$. Note that, even if such a $\mathcal{J}$ interprets b and, say, c as the same object, still, we can construct a model of $\mathcal{K}_1$ which interprets $aw_{[R]}$ and $aw_{[S]}$ as different objects: namely, $aw_{[R]}^{\mathcal{I}} = b^{\mathcal{J}}$ and $aw_{[S]}^{\mathcal{I}} = d^{\mathcal{J}}$.

What goes against our intuition obtained from the previous development, is that there is *no* inequality between b and c, nor between b and d, but still, $\mathcal{A}_2$ is a universal solution for $\mathcal{K}_1$ under $\mathcal{M}$. This example suggests that one needs to reason about the cardinality of the domain, moreover, homomorphisms, even extended ones, cannot be used to characterize universal solutions when we allow for inequalities between constants. ■

We conclude with some directions for future work.

Many of the existing results can be extended to other Horn DLs, such as *DL-Lite*$^{\mathcal{H}}_{horn}$, $\mathcal{ELH}$, and Horn-$\mathcal{ALCHI}$, namely the algorithms based on reachability games and two-way alternating automata, both of which heavily rely on the notion of the canonical model.

In this thesis we have not dealt with other "standard" data exchange reasoning tasks, such as composition and inversion of mappings [53, 49, 54, 4]. These problems are certainly of interest in the KB exchange framework.

The work presented in this thesis is assumed to lay foundational basis for KB exchange and contains only purely theoretical results. So one of the next steps is to see how the obtained results can be applied to practice. This would require to develop a prototype system for KB exchange, in which the devised algorithms could be implemented.

Part II

APPENDIX

A
APPENDIX

A.1 THE THEORY OF AUTOMATA, LOGIC, AND INFINITE GAMES

In this section we define two important theories historically emerged in an attempt to construct and verify reactive programs, such as communication protocols or control systems: alternating tree automata and infinite two-person games. While their primary application is that of model-checking, in this thesis we use both of them for checking homomorphism between trees generated by *DL-Lite*$_{\mathcal{R}}$ KBs. We are employing a game-theoretic approach in Section 5.2 for deciding the membership problem for universal solutions with simple ABoxes, and an automata-theoretic approach in Section 5.3.2 for deciding the non-emptiness problem for universal solutions with extended ABoxes.

A.1.1 *Two-way Alternating Tree Automata*

Alternating automata on infinite trees are a generalization of nondeterministic automata on infinite trees, introduced in [92], which are in turn a generalization of automata on infinite words, introduced by Büchi in [26]. They allow for an elegant reduction of decision problems for temporal and program logics [46, 21], and μ-calculus [45, 100, 103], which are traditionally used in the automatic verification and synthesis of hardware and software systems. In this section we are going to define two-way alternating tree automata introduced in [102] that are alternating automata specifically design to handle backward modalities of μ-calculus, and thus, perfectly suited to deal with inverse roles in *DL-Lite*$_{\mathcal{R}}$.

We start with giving some necessary definitions.Infinite trees are represented as prefix closed (infinite) sets of words over $\mathbb{N}$ (the set of positive natural numbers). Formally, an *infinite tree* is a set of words $T \subseteq \mathbb{N}^*$, such that if $x \cdot c \in T$, where $x \in \mathbb{N}^*$ and $c \in \mathbb{N}$, then also $x \in T$. The elements of T are called nodes, the empty word ϵ is the root of T, and for every $x \in T$, the nodes $x \cdot c$, with $c \in \mathbb{N}$, are the successors of x. By convention we take $x \cdot 0 = x$, and $x \cdot i \cdot -1 = x$. The branching degree $d(x)$ of a node x denotes the number of successors of x. If the branching degree of all nodes of a tree is bounded by k, we say that the tree has branching degree k. An infinite path P of T is a prefix closed set $P \subseteq T$ such that for every $i \geq 0$ there exists a unique node $x \in P$ with $|x| = i$. A labeled tree over an alphabet Σ is a pair (T, V), where T is a tree and $V : T \to \Sigma$ maps each node of T to an element of Σ.

In alternating automata, transition function is a Boolean formula. Therefore, let $\mathcal{B}(I)$ be the set of positive Boolean formulae over I, built inductively by applying $\wedge$ and $\vee$ starting from true, false, and elements of I. For a set $J \subseteq I$ and a formula $\phi \in \mathcal{B}(I)$, we say that J satisfies ϕ if and only if, assigning true to the elements in J and false to those in $I \setminus J$, makes ϕ true. For a positive integer k, let $[k] = \{-1, 0, 1, \ldots, k\}$.

Finally, a *two-way alternating tree automaton (2ATA)* running over infinite trees with branching degree k, is a tuple $\mathbb{A} = \langle \Sigma, Q, \delta, q_0, F \rangle$, where Σ is the input alphabet, Q is a finite set of states, $\delta : Q \times \Sigma \to \mathcal{B}([k] \times Q)$ is the transition function, $q_0 \in Q$ is the initial state, and F specifies the acceptance condition.

The transition function maps a state $q \in Q$ and an input letter $\sigma \in \Sigma$ to a positive boolean formula over $[k] \times Q$. Intuitively, if $\delta(q, \sigma) = \phi$, then each pair (c, q') appearing in ϕ corresponds to a new copy of the automaton going to the direction suggested by c and starting in state q'. For example, if $k = 2$ and $\delta(q_1, \sigma) = ((1, q_2) \wedge (1, q_3)) \vee ((-1, q_1) \wedge (0, q_3))$, when the automaton is in the state q_1 and is reading the node x labeled by the letter σ, it proceeds either by sending off two copies, in the states q_2 and q_3 respectively, to the first successor of x (i.e., $x \cdot 1$), or by sending off one copy in the state q_1 to the predecessor of x (i.e., $x \cdot -1$) and one copy in the state q_3 to x itself (i.e., $x \cdot 0$).

A *run* of a 2ATA $\mathbb{A}$ over a labeled tree (T, V) is a labeled tree $(T_{\mathbf{r}}, \mathbf{r})$ in which every node is labeled by an element of $T \times Q$. A node in $T_{\mathbf{r}}$ labeled by (x, q) describes a copy of A that is in the state q and reads the node x of T. The labels of adjacent nodes have to satisfy the transition function of $\mathbb{A}$. Formally, a *run* $(T_{\mathbf{r}}, \mathbf{r})$ is a $T \times Q$-labeled tree satisfying:

- $\epsilon \in T_{\mathbf{r}}$ and $\mathbf{r}(\epsilon) = (\epsilon, q_0)$.
- Let $y \in T_{\mathbf{r}}$, with $\mathbf{r}(y) = (x, q)$ and $\delta(q, V(x)) = \phi$. Then there is a (possibly empty) set $S = \{(c_1, q_1), \ldots, (c_n, q_n)\} \subseteq [k] \times Q$ such that:
  - S satisfies ϕ and
  - for all $1 \leq i \leq n$, we have that $y \cdot i \in T_{\mathbf{r}}$, $x \cdot c_i$ is defined ($x \cdot c_i \in T$), and $\mathbf{r}(y \cdot i) = (x \cdot c_i, q_i)$.

A run $(T_{\mathbf{r}}, \mathbf{r})$ is accepting if *all* its infinite paths satisfy the acceptance condition. Given an infinite path $P \in T_{\mathbf{r}}$, let $\mathit{inf}(P) \subseteq Q$ be the set of states that appear infinitely often in P (as second components of node labels). We consider here Büchi acceptance conditions [26]. A Büchi condition over a state set Q is a subset F of Q, and an infinite path P satisfies F if $\mathit{inf}(P) \cap F \neq \emptyset$. Notice that if a run does not contain infinite paths, then it trivially satisfies the acceptance condition.

The non-emptiness problem for 2ATAs consists in determining, for a given 2ATA, whether the set of trees it accepts is nonempty. It is known that this problem can be solved in exponential time in the number of states of the input automaton $\mathbb{A}$, but in linear time in the size of the alphabet as well as in the size of the transition function of $\mathbb{A}$ [102]. Moreover, a tree (T, V) is in the language $\mathcal{L}(\mathbb{A})$ of a 2ATA $\mathbb{A}$ if and only if there exists an accepting run of $\mathbb{A}$ over (T, V).

A.1.2 *Reachability games on graphs*

Infinite two-person games on directed graphs provide tools for nice and intuitive proofs for logics over trees, and are closely related to automata on infinite words, including the two-way alternating tree automata defined in the previous section: the non-emptiness problem for 2ATA is solved using infinite games. The idea of infinite games arose implicitly in the study of synthesis of digital circuits in the 1960s; McNaughton was the first to explicitly use the term "infinite games" [89, 90] already in 1965. For a good introduction to infinite games the reader can refer to [87].

The accepting conditions for games, which specify when a particular play is a win for Player 0, are the same as in the case of automata. In this thesis we will consider reachability acceptance conditions, a simpler case of Büchi acceptance conditions, hence in this section we define reachability games.

A game is defined by a game graph (a playground) and a winning condition.

A *game graph* is a triple $\mathsf{G} = (\mathsf{S}_0, \mathsf{S}_1, \mathsf{T})$, where $\mathsf{S} = \mathsf{S}_0 \cup \mathsf{S}_1$ is a finite set of states, $\mathsf{S}_0 \cap \mathsf{S}_1 = \emptyset$ and $\mathsf{T} \subseteq \mathsf{S} \times \mathsf{S}$ is a transition relation. A play in such a game graph can be seen as moving a pebble from one state to another via transition edges starting from some initial state. The game starts in some state $s_0 \in \mathsf{S}$, and it is played in turns. In each turn, if the current state s is in S_i ($i = 0, 1$), then Player i chooses some state $s' \in \mathsf{S}$ such that $(s, s') \in \mathsf{T}$. This is repeated either infinitely often or until a state without successors, a dead end, is reached. Formally, a state s is called a *dead end* if $\{s' \mid (s, s') \in \mathsf{T}\} = \emptyset$, and a *play* in a game graph G can be

- an infinite path π in G, $\pi = s_0 s_1 s_2 \ldots$ such that $s_i \in \mathsf{S}$ and $(s_i, s_{i+1}) \in \mathsf{T}$ for each $i \geq 0$, then it is called an *infinite play*,

- a finite path $\pi = s_0 s_1 \ldots s_k \in \mathsf{S}^{k+1}$ such that $(s_i, s_{i+1}) \in \mathsf{T}$ for every $i \in \{0, \ldots, k-1\}$, and s_k is a dead end, then π is called a *finite play*.

The *winning condition* defines what are the plays won by Player 0. We will consider a *reachability acceptance condition* specified as follows: given a set of accepting states $F \subseteq \mathsf{S}$, a play π is a *win* for Player 0 iff some vertex from F occurs in π. A *strategy for Player 0 from* state s is a (partial) function $f_0 : \mathsf{S}^\star \mathsf{S}_0 \to \mathsf{S}$ such that it assigns to each sequence of states $s_0, s_1, \ldots, s_k$ with $s_0 = s$ and $s_k \in \mathsf{S}_0$, a successor state s_{k+1} such that $(s_k, s_{k+1}) \in \mathsf{T}$. A play $\pi = s_0 s_1 \cdots$ is said to *conform* with strategy f_0 if $s_{i+1} = f_0(s_0 s_1 \ldots s_i)$ for every $s_i \in \mathsf{S}_0$. Then, a strategy f_0 is a *winning strategy* for Player 0 *from* $s \in \mathsf{S}$ if every play that conforms with f_0 and starts in s is a win for Player 0, and Player 0 *wins from* s if he has a winning strategy from s. Finally, the *winning region* of Player 0, denoted $W_0 \subseteq \mathsf{S}$ is the set of all states s such that he wins from s. The corresponding notions for Player 1 are defined analogously.

Finally, a *reachability game* is a pair $G = (\mathsf{G}, F)$ where G is a game graph and F is a set of accepting states.

Proposition A.1.1 ([87],[33]). *Given a reachability game $G = (\mathsf{G}, F)$ and a state s in G, it can be checked in* PTIME *whether Player 0 has a winning strategy from s.*

BIBLIOGRAPHY

[1] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison Wesley Publ. Co., 1995.

[2] P. Adjiman, P. Chatalic, F. Gouasdoué, M.-C. Rousset, and L. Simon. Distributed reasoning in a peer-to-peer setting: Application to the semantic web. *J. of Artificial Intelligence Research*, 25:269–314, 2006.

[3] José-Luis Aguirre, Kai Eckert, Jérôme Euzenat, Alfio Ferrara, Willem Robert van Hage, Laura Hollink, Christian Meilicke, Andriy Nikolov, Dominique Ritze, François Scharffe, Pavel Shvaiko, Ondrej Sváb-Zamazal, Cássia Trojahn dos Santos, Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, and Benjamin Zapilko. Results of the ontology alignment evaluation initiative 2012. In *Proc. of the 7th International Workshop on Ontology Matching, Boston, MA, USA, November 11, 2012*, 2012.

[4] Marcelo Arenas, Jorge Pérez, and Cristian Riveros. The recovery of a schema mapping: Bringing exchanged data back. *ACM Trans. Database Syst.*, 34:22:1–22:48, December 2009. ISSN 0362-5915. doi: http://doi.acm.org/10.1145/1620585.1620589. URL http://doi.acm.org/10.1145/1620585.1620589.

[5] Marcelo Arenas, Elena Botoeva, and Diego Calvanese. Knowledge base exchange. In *Proc. of the 24th Int. Workshop on Description Logic (DL 2011)*, volume 745 of *CEUR Electronic Workshop Proceedings,* http://ceur-ws.org/, 2011.

[6] Marcelo Arenas, Jorge Pérez, and Juan L. Reutter. Data exchange beyond complete data. In *Proc. of the 30th ACM SIGMOD SIGACT SIGART Symp. on Principles of Database Systems (PODS 2011)*, pages 83–94, 2011.

[7] Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Exchanging description logic knowledge bases. In *Proc. of the 13th Int. Conf. on Knowledge Representation and Reasoning (KR 2012)*, pages 563–567. AAAI Press, 2012.

[8] Marcelo Arenas, Elena Botoeva, Diego Calvanese, Vladislav Ryzhikov, and Evgeny Sherkhonov. Representability in *DL-Lite*$_r$ knowledge base exchange. In *Proc. of the 25th Int. Workshop on Description Logic (DL 2012)*, volume 846 of *CEUR Electronic Workshop Proceedings,* http://ceur-ws.org/, 2012.

[9] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Exchanging OWL 2 QL knowledge bases. In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI 2013)*, pages 703–710, 2013.

[10] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Exchanging OWL 2 QL knowledge bases (extended version). CoRR Technical Report arXiv:1304.5810, arXiv.org e-Print archive, 2013. Available at http://arxiv.org/abs/1304.5810.

[11] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Computing solutions in OWL 2 QL knowledge exchange. In *Proc. of the 26th Int. Workshop on Description Logic (DL 2013)*, volume 1014 of *CEUR Electronic Workshop Proceedings*, http://ceur-ws.org/, pages 4–16, 2013.

[12] Marcelo Arenas, Jorge Pérez, and Juan Reutter. Data exchange beyond complete data. *J. of the ACM*, 60(4):28:1–28:59, September 2013. ISSN 0004-5411. doi: 10.1145/2505985. URL http://doi.acm.org/10.1145/2505985.

[13] Marcelo Arenas, Jorge Pérez, and Emanuel Sallinger. Towards general representability in knowledge exchange. In *Proc. of the 7th Alberto Mendelzon Int. Workshop on Foundations of Data Management (AMW 2013)*, 2013.

[14] Marcelo Arenas, Elena Botoeva, Diego Calvanese, and Vladislav Ryzhikov. Knowledge base exchange: The case of OWL 2 QL. Under submission to an international journal, 2014.

[15] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The *DL-Lite* family and relations. *J. of Artificial Intelligence Research*, 36:1–69, 2009.

[16] Franz Baader. Terminological cycles in a description logic with existential restrictions. In *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003)*, pages 325–330, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

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

[18] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005)*, pages 364–369, 2005.

[19] Jie Bao et al. OWL 2 Web Ontology Language document overview (second edition). W3C Recommendation, World Wide Web Consortium, December 2012. http://www.w3.org/TR/owl2-overview/.

[20] Pablo Barceló. Logical foundations of relational data exchange. *SIGMOD Record*, 38(1):49–58, 2009.

[21] Orna Bernholtz, Moshe Y. Vardi, and Pierre Wolper. An automata-theoretic approach to branching-time model checking. In *Proc. of the 6th Int. Conf. on Computer Aided Verification (CAV'94)*, volume 818 of *Lecture Notes in Computer Science*, pages 142–155. Springer, 1994.

[22] Egon Börger, Erich Grädel, and Yuri Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997.

[23] Elena Botoeva, Roman Kontchakov, Vladislav Ryzhikov, Frank Wolter, and Michael Zakharyaschev. Query inseparability for description logic knowledge bases. In *Proc. of the 14th Int. Conf. on Knowledge Representation and Reasoning (KR 2014)*. AAAI Press, 2014.

[24] Ronald J. Brachman and Hector J. Levesque, editors. *Readings in Knowledge Representation*. Morgan Kaufmann, 1985.

[25] Dan Brickley and R. V. Guha. RDF vocabulary description language 1.0: RDF Schema. W3C Recommendation, February 2004. Available at http://www.w3.org/TR/rdf-schema/.

[26] J. Richard Büchi. On a decision method in restricted second order arithmetic. In E. Nagel et al., editors, *Proc. Internat. Congr. on Logic, Methodology and Philosophy of Science*, pages 1–11. Stanford University Press, 1960.

[27] Sasa Buvac and Ian A. Mason. Propositional logic of context. In *Proc. of the 11th National Conf. on Artificial Intelligence (AAAI 1993)*, pages 412–419, 1993.

[28] Saša Buvač and Richard Fikes. A declarative formalization of knowledge translation. In *Proc. of the 4th Int. Conf. on Information and Knowledge Management (CIKM 1995*, pages 340–347, New York, NY, USA, 1995. ACM. ISBN 0-89791-812-6. doi: 10.1145/221270.221610. URL http://doi.acm.org/10.1145/221270.221610.

[29] Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. 2ATAs make DLs easy. In *Proc. of the 15th Int. Workshop on Description Logic (DL 2002)*, volume 53 of *CEUR Electronic Workshop Proceedings, http://ceur-ws.org/*, pages 107–118, 2002.

[30] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.

[31] Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Answering regular path queries in expressive description logics: An automata-theoretic approach. In *Proc. of the 22nd Nat. Conf. on Artificial Intelligence (AAAI 2007)*, pages 391–396, 2007.

[32] Diego Calvanese, Domenico Carbotta, and Magdalena Ortiz. A practical automata-based technique for reasoning in expressive description logics. In *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI 2011)*, pages 798–804, 2011.

[33] Krishnendu Chatterjee and Monika Henzinger. An o(n2) time algorithm for alternating Büchi games. In *Proc. of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2012)*, pages 1386–1399, 2012. URL http://dl.acm.org/citation.cfm?id=2095116.2095225.

[34] Namyoun Choi, Il-Yeol Song, and Hyoil Han. A survey on ontology mapping. *SIGMOD Rec.*, 35(3):34–41, September 2006. ISSN 0163-5808.

doi: 10.1145/1168092.1168097. URL http://doi.acm.org/10.1145/1168092.1168097.

[35] R Cote, D Rothwell, J Palotay, R Beckett, and L Brochu. The systematized nomenclature of human and veterinary medicine. Technical report, SNOMED International, Northfield, IL: College of American Pathologists, 1993.

[36] Bernardo Cuenca Grau, Ian Horrocks, Oliver Kutz, and Ulrike Sattler. Will my ontologies fit together? In *Proc. of the 19th Int. Workshop on Description Logic (DL 2006)*, volume 189 of *CEUR Electronic Workshop Proceedings, http://ceur-ws.org/*, 2006.

[37] Bernardo Cuenca Grau, Bijan Parsia, Evren Sirin, and Aditya Kalyanpur. Modularity and Web ontologies. In *Proc. of the 10th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2006)*, pages 198–209, 2006.

[38] Bernardo Cuenca Grau, Ian Horrocks, Yevgeny Kazakov, and Ulrike Sattler. A logical framework for modularity of ontologies. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI 2007)*, 2007.

[39] Răzvan Diaconescu, Joseph Goguen, and Petros Stefaneas. Logical support for modularisation. In *Papers Presented at the Second Annual Workshop on Logical Environments*, pages 83–130, New York, NY, USA, 1993. Cambridge University Press. ISBN 0-521-43312-6. URL http://dl.acm.org/citation.cfm?id=185881.185899.

[40] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Werner Nutt. The complexity of concept languages. In *Proc. of the 2nd Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'91)*, pages 151–162, 1991.

[41] Dejing Dou and Drew Mcdermott. Deriving axioms across ontologies. In *Proc. of Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS 06)*, pages 952–954, 2006.

[42] Dejing Dou, Drew Mcdermott, and Peishen Qi. Ontology translation on the semantic web. In *Journal of Data Semantics*, pages 35–57, 2004.

[43] Dejing Dou, Han Qin, and Haishan Liu. Semantic translation for rule-based knowledge in data mining. In *Proc. of the 22nd Int. Conf. on Database and Expert Systems Applications - Volume Part II (DEXA 2011)*, pages 74–89, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-23090-5. URL http://dl.acm.org/citation.cfm?id=2033546.2033555.

[44] Thomas Eiter, Georg Gottlob, Magdalena Ortiz, and Mantas Šimkus. Query answering in the description logic Horn-$\mathcal{SHIQ}$. In *Proc. of the 11th Eur. Conference on Logics in Artificial Intelligence (JELIA 2008)*, pages 166–179, 2008.

[45] E. Allen Emerson and Charanjit S. Jutla. The complexity of tree automata and logics of programs. In *Proc. of the 29th Annual Symp. on the Foundations of Computer Science (FOCS'88)*, pages 328–337, 1988.

[46] E. Allen Emerson and Charanjit S. Jutla. Tree automata, mu-calculus and determinacy. In *Proc. of the 32nd Annual Symp. on the Foundations of Computer Science (FOCS'91)*, pages 368–377, 1991.

[47] Jérôme Euzenat and Pavel Schwaiko. *Ontology Matching*. Springer, 2007.

[48] Jérôme Euzenat, Christian Meilicke, Heiner Stuckenschmidt, Pavel Shvaiko, and Cássia Trojahn dos Santos. Ontology alignment evaluation initiative: Six years of experience. *Journal on Data Semantics XV*, 15:158–192, 2011.

[49] Ronald Fagin. Inverting schema mappings. *ACM Trans. on Database Systems*, 32(4), 2007.

[50] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. In *Proc. of the 9th Int. Conf. on Database Theory (ICDT 2003)*, pages 207–224, 2003.

[51] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. *Theoretical Computer Science*, 336(1): 89–124, 2005.

[52] Ronald Fagin, Phokion G. Kolaitis, and Lucian Popa. Data exchange: Getting to the core. *ACM Trans. on Database Systems*, 30(1):174–210, 2005.

[53] Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang-Chiew Tan. Composing schema mappings: Second-order dependencies to the rescue. *ACM Trans. on Database Systems*, 30(4):994–1055, 2005.

[54] Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang-Chiew Tan. Quasi-inverses of schema mappings. *ACM Trans. on Database Systems*, 33(2):1–52, 2008.

[55] Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang Chiew Tan. Reverse data exchange: coping with nulls. In *Proc. of the 28th ACM SIGMOD SIGACT SIGART Symp. on Principles of Database Systems (PODS 2009)*, pages 23–32, 2009.

[56] M. Fikes, R.and Cutkosky, T. R. Gruber, and J. Van Baalen. Knowledge sharing technology project overview. Technical Report KSL 91-71, Knowledge Systems Laboratory, Stanford University, 1991.

[57] A. Fuxman, P. G. Kolaitis, R. J. Miller, and W. C. Tan. Peer data exchange. *ACM Trans. on Database Systems*, 31(4):1454–1498, 2005.

[58] M. R. Garey and D. S. Johnson. *Computers and Intractability — A guide to NP-completeness*. W. H. Freeman and Company, San Francisco (CA, USA), 1979.

[59] Silvio Ghilardi, Carsten Lutz, and Frank Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In *Proc. of the 10th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2006)*, pages 187–197, 2006.

[60] Fausto Giunchiglia, Mikalai Yatskevich, Paolo Avesani, and Pavel Shvaiko. A large dataset for the evaluation of ontology matching. *Knowledge Eng. Review*, 24(2):137–157, 2009.

[61] Georg Gottlob and Alan Nash. Data exchange: Computing cores in polynomial time. In *Proc. of the 25th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2006)*, pages 40–49, 2006.

[62] Georg Gottlob and Alan Nash. Efficient core computation in data exchange. *J. ACM*, 55(2):9:1–9:49, May 2008. ISSN 0004-5411. doi: 10.1145/1346330.1346334. URL http://doi.acm.org/10.1145/1346330.1346334.

[63] Erich Grädel and Igor Walukiewicz. Guarded fixed point logic. In *Proc. of the 14th IEEE Symp. on Logic in Computer Science (LICS'99)*, pages 45–54. IEEE Computer Society Press, 1999.

[64] Ramanathan Guha. *Contexts: A Formalization and Some Applications*. PhD thesis, Stanford University, Stanford, CA, USA, 1992. UMI Order No. GAX92-17827.

[65] Laura M. Haas, Mauricio A. Hernández, Howard Ho, Lucian Popa, and Mary Roth. Clio grows up: from research prototype to industrial tool. In *Proc. of the ACM SIGMOD Int. Conf. on Management of data (SIGMOD 2005)*, pages 805–810, New York, NY, USA, 2005. ACM. ISBN 1-59593-060-4. URL http://doi.acm.org/10.1145/1066157.1066252.

[66] André Hernich. Computing universal models under guarded tgds. In *Proc. of the 15th Int. Conf. on Database Theory (ICDT 2012)*, pages 222–235. ACM, 2012.

[67] André Hernich and Nicole Schweikardt. Logic and data exchange: Which solutions are "good" solutions? In *Proc. of the 8th Int. Conf. on Logic and the Foundations of Game and Decision Theory (LOFT 2008)*, pages 61–85, 2008.

[68] Ian Horrocks and Ulrike Sattler. A tableaux decision procedure for $\mathcal{SHOIQ}$. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005)*, pages 448–453, 2005.

[69] Ian Horrocks, Ulrike Sattler, and Stephan Tobies. Practical reasoning for very expressive description logics. *J. of the Interest Group in Pure and Applied Logic*, 8(3):239–264, 2000.

[70] Wei Hu, Yuzhong Qu, and Gong Cheng. Matching large ontologies: A divide-and-conquer approach. *Data Knowl. Eng.*, 67(1):140–160, 2008.

[71] Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Yujiao Zhou, and Ian Horrocks. Large-scale interactive ontology matching: Algorithms and implementation. In *Proc. of the 20th European Conf. on Artificial Intelligence (ECAI 2012)*, pages 444–449, 2012.

[72] David S. Johnson and Anthony C. Klug. Testing containment of conjunctive queries under functional and inclusion dependencies. *J. of Computer and System Sciences*, 28(1):167–189, 1984.

[73] Charanjit S. Jutla. Determinization and memoryless winning strategies. *Inf. Comput.*, 133(2):117–134, March 1997. ISSN 0890-5401. doi: 10.1006/inco.1997.2624. URL http://dx.doi.org/10.1006/inco.1997.2624.

[74] Anastasios Kementsietsidis, Marcelo Arenas, and Renée J. Miller. Mapping data in peer-to-peer systems: Semantics and algorithmic issues. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data (SIGMOD 2003)*, pages 325–336, 2003.

[75] Phokion G. Kolaitis. Schema mappings, data exchange, and metadata management. In *Proc. of the 24th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2005)*, pages 61–75, 2005.

[76] Boris Konev, Carsten Lutz, Dirk Walther, and Frank Wolter. Semantic modularity and module extraction in description logics. In *Proc. of the 18th European Conf. on Artificial Intelligence (ECAI 2008)*, pages 55–59. IOS Press, 2008.

[77] Boris Konev, Roman Kontchakov, Michel Ludwig, Thomas Schneider, Frank Wolter, and Michael Zakharyaschev. Conjunctive query inseparability of OWL 2 QL TBoxes. In *Proc. of the 26th Nat. Conf. on Artificial Intelligence (AAAI 2011)*, 2011.

[78] Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev. Can you tell the difference between *DL-Lite* ontologies? In *Proc. of the 11th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2008)*, pages 285–295, 2008.

[79] Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev. Logic-based ontology comparison and module extraction, with an application to dl-lite. *Artificial Intelligence*, 174(15):1093–1141, 2010.

[80] Dexter Kozen. Results on the propositional μ-calculus. *Theoretical Computer Science*, 27:333–354, 1983.

[81] Dexter Kozen. *Theory of Computation*. Springer, 2006.

[82] Patrick Lambrix and He Tan. Sambo - a system for aligning and merging biomedical ontologies. *J. of Web Semantics*, 4(3):196–206, 2006.

[83] Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proc. of the 21st ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems (PODS 2002)*, pages 233–246, 2002.

[84] Leonid Libkin and Cristina Sirangelo. Data exchange and schema mappings in open and closed worlds. *Journal of Computer and System Sciences*, 77(3):542–571, 2011.

[85] Carsten Lutz and Frank Wolter. Conservative extensions in the lightweight description logic $\mathcal{EL}$. In *Proc. of the 21st Int. Conf. on Automated Deduction (CADE 2007)*, 2007.

[86] Carsten Lutz, Dirk Walther, and Frank Wolter. Conservative extensions in expressive description logics. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI 2007)*, pages 453–458, 2007.

[87] René Mazala. Infinite games. In *Automata, Logics, and Infinite Games*, pages 23–42, 2001.

[88] John McCarthy. Notes on formalizing context. In *Proc. of the 13th Int. Joint Conf. on Artifical Intelligence - Volume 1 (IJCAI'93)*, pages 555–560, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc. URL http://dl.acm.org/citation.cfm?id=1624025.1624103.

[89] Robert McNaughton. Finite-state infinite games. Technical report, Project MAC, Massachusetts Institute of Technology, USA, 1965.

[90] Robert McNaughton. Infinite games played on finite graphs. *Annals of Pure and Applied Logic*, 65(2):149–184, 1993.

[91] Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz. OWL 2 Web Ontology Language profiles (second edition). W3C Recommendation, World Wide Web Consortium, December 2012. http://www.w3.org/TR/owl2-profiles/.

[92] D. E. Muller and P. E. Schupp. Alternating automata on infinite trees. *Theoretical Computer Science*, 54:267–276, 1987.

[93] David E. Muller and Paul E. Schupp. The theory of ends, pushdown automata, and second-order logic. *Theoretical Computer Science*, 37:51–75, 1985.

[94] Natalya F. Noy and Mark A. Musen. The prompt suite: Interactive tools for ontology merging and mapping. *Int. J. of Human-Computer Studies*, 59(6):983–1024, December 2003. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2003.08.002. URL http://dx.doi.org/10.1016/j.ijhcs.2003.08.002.

[95] Christos H. Papadimitriou. *Computational Complexity*. Addison Wesley Publ. Co., 1994.

[96] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *J. on Data Semantics*, X:133–173, 2008.

[97] Stefan Schulz, Kornel Marko, and Boontawee Suntisrivaraporn. Formal representation of complex SNOMED CT expressions. *BMC Medical Informatics and Decision Making*, 8(1), 2008. doi: 10.1186/1472-6947-8-S1-S9. URL http://www.biomedcentral.com/1472-6947/8/S1/S9.

[98] N. C. Shu, B. C. Housel, R. W. Taylor, S. P. Ghosh, and V. Y. Lum. Express: a data extraction, processing, and restructuring system. *ACM Trans. Database Syst.*, 2:134–174, June 1977. ISSN 0362-5915. doi: http://doi.acm.org/10.1145/320544.320549. URL http://doi.acm.org/10.1145/320544.320549.

[99] Pavel Shvaiko and Jérôme Euzenat. Ontology matching: State of the art and future challenges. *IEEE Trans. on Knowledge and Data Engineering*, 25(1):158–176, 2013.

[100] Robert S. Streett. Propositional Dynamic Logic of looping and converse is elementarily decidable. *Information and Control*, 54:121–141, 1982.

[101] Heiner Stuckenschmidt, Christine Parent, and Stefano Spaccapietra, editors. *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, volume 5445 of *Lecture Notes in Computer Science*. Springer, 2009. ISBN 978-3-642-01906-7.

[102] Moshe Y. Vardi. Reasoning about the past with two-way automata. In *Proc. of the 25th Int. Coll. on Automata, Languages and Programming (ICALP'98)*, volume 1443 of *Lecture Notes in Computer Science*, pages 628–641. Springer, 1998.

[103] Moshe Y. Vardi and Pierre Wolper. Automata-theoretic techniques for modal logics of programs. *J. of Computer and System Sciences*, 32:183–221, 1986.

[104] Heribert Vollmer. *Introduction to Circuit Complexity: A Uniform Approach*. Springer, 1999.

[105] Igor Walukiewicz. Pushdown processes: Games and model-checking. *Information and Computation*, 164(2):234–263, 2001.