Unification in ELH$_{R^+}$ without the Top Concept modulo Cycle-Restricted Ontologies (Extended Abstract)

Franz Baader[1,2], Oliver Fernández Gil[1,2]

[1]*Theoretical Computer Science, TU Dresden, Germany*

[2]*Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany*

Abstract

Unification has been introduced in Description Logic (DL) as a means to detect redundancies in ontologies. In particular, it was shown that testing unifiability in the DL $\mathcal{EL}$ is an NP-complete problem, and this result has been extended in several directions. Surprisingly, it turned out that the complexity increases to PSpace if one disallows the use of the top concept in concept descriptions. Motivated by features of the medical ontology SNOMED CT, we extend this result to a setting where the top concept is disallowed, but there is a background ontology consisting of restricted forms of concept and role inclusion axioms. We are able to show that the presence of such axioms does not increase the complexity of unification without top, i.e., testing for unifiability remains a PSpace-complete problem.

Keywords

Unification, Description Logics, Complexity

Motivation for unification in DLs

Unification in Description Logics (DLs) has been introduced in [1] as a new inference service, motivated by the need for detecting redundancies in ontologies, in a setting where different ontology engineers (OEs) constructing the ontology may model the same concepts on different levels of granularity. For example, assume that (using the style of definitions in the medical ontology SNOMED CT[1]) one OE models the concept of a *viral infection of the lung* as

$$ViralInfection \sqcap \exists findingSite.LungStructure,$$

whereas another one models it as

$$LungInfection \sqcap \exists causativeAgent.Virus.$$

Here *ViralInfection* and *LungInfection* are used as atomic concepts without further defining them, i.e., the two OEs made different decisions when to stop the modelling

[1]https://www.ihtsdo.org/snomed-ct/

process. The resulting concept descriptions are not equivalent, but they are nevertheless meant to represent the same concept. They can be made equivalent by treating the concept names *ViralInfection* and *LungInfection* as variables, and then substituting the first one by *Infection* $\sqcap \exists causativeAgent.Virus$ and the second one by *Infection* $\sqcap \exists findingSite.LungStructure$. In this case, we say that the descriptions are unifiable, and call the substitution that makes them equivalent a *unifier*. Intuitively, such a unifier proposes definitions for the concept names that are used as variables. In [2], unification and its extension to disunification are used to construct new medical concepts from SNOMED CT.

Unification in the DLs $\mathcal{EL}$ and $\mathcal{EL}^{-\top}$

Unification in the DL $\mathcal{EL}$ (which provides the concept constructors conjunction $\sqcap$, existential restriction $\exists r.C$, and top concept $\top$) was first investigated in [3], where it was proved that deciding unifiability is an NP-complete problem. The NP upper bound was shown in that paper using a brute-force "guess and then test" NP algorithm. More practical algorithms for solving this problem and for computing unifiers were presented in [4] and [5], where the former describes a goal-oriented transformation-based algorithm and the latter is based on a translation to SAT. Implementations of these two algorithms are provided by the system UEL[2] [6], which is also available as a plug-in for the ontology editor Protégé. At the time these algorithms were developed, SNOMED CT was an $\mathcal{EL}$ ontology consisting of acyclic concept definitions. Since such definitions can be encoded into the unification problem (see Section 2.3 in [4]), algorithms for unification of $\mathcal{EL}$ concept descriptions (without background ontology) could be applied to SNOMED CT.

There was, however, one problem with employing these algorithms in the context of SNOMED CT: the top concept is not used in SNOMED CT, but the concepts generated by $\mathcal{EL}$ unification might contain $\top$, even if applied to concept descriptions not containing $\top$. Thus, the concept descriptions produced by the unifier are not necessarily in the style of SNOMED CT. For example, assume that we are looking for a unifier satisfying the two subsumption constraints[3]

$$\exists findingSite.LungStructure \sqsubseteq^? \exists findingSite.X,$$

$$\exists findingSite.HeartStructure \sqsubseteq^? \exists findingSite.X.$$

It is easy to see that there is only one unifier of these two constraints, which replaces X with $\top$. Unification in $\mathcal{EL}^{-\top}$, i.e., the fragment of $\mathcal{EL}$ in which the top constructor is disallowed, was investigated in [8, 9]. Surprisingly, it turned out that the absence of $\top$ makes unification considerably harder, both from a conceptual and a computational complexity point of view. In fact, the complexity of deciding unifiability increases from NP-complete for $\mathcal{EL}$ to PSpace-complete for $\mathcal{EL}^{-\top}$. The unification algorithm for $\mathcal{EL}^{-\top}$

---

[2]https://sourceforge.net/projects/uel/

[3]Instead of equivalence constraints, as in our above example and in early work on unification in DLs, we consider here a set of subsumption constraints as unification problem. It is easy to see that these two kinds of unification problems can be reduced to each other [7].

introduced in [8, 9] basically proceeds as follows. It first applies the unification algorithm for $\mathcal{EL}$ to compute so-called local unifiers. If none of them is an $\mathcal{EL}^{-\top}$-unifier, then it tries to augment the images of the variables by conjoining a special kind of concept descriptions called *particles*. The task of finding appropriate particles is reduced to solving certain systems of linear language inclusions, which can be realized in PSpace using an automata-based approach.

Extending unification in $\mathcal{EL}$ towards background ontologies

The current version of SNOMED CT consists not only of acyclic concept definitions, but also contains more general concept inclusions (GCIs). In addition, properties of the part-of relation are no longer encoded using the so-called SEP-triplet encoding [10], but are directly expressed via role axioms [11], which can, for instance, be used to state that the part-of relation is transitive and that proper-part-of is a subrole of part-of.

Decidability of unification in $\mathcal{EL}$ w.r.t. a background ontology consisting of GCIs is still an open problem. In [7], it is shown that the problem remains in NP if the ontology is cycle-restricted, which is a condition that the current version of SNOMED CT satisfies. Extensions of this result to the DL $\mathcal{ELH}_{\mathcal{R}+}$, which additionally allows for transitive roles and role inclusion axioms, were presented in [12] and [13], where the former introduces a SAT-based algorithm and the latter a transformation-based one. However, in all these algorithms, unifiers may introduce concept descriptions containing $\top$. In our example with the different finding site, however, the presence of the GCIs

$$LungStructure \sqsubseteq UpperBodyStructure \quad \text{and} \quad HeartStructure \sqsubseteq UpperBodyStructure$$

would yield a unifier not using $\top$, namely the one that replaces X with *UpperBodyStructure*.

Our new contribution

The purpose of the work we report on here is to investigate the unification problem in the DL $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$ w.r.t. cycle-restricted ontologies. Our main result is that the presence of a cycle-restricted ontology does not increase the complexity of unification in $\mathcal{EL}$ without top, i.e., the unification decision problem remains in PSpace.

Theorem 1. *Deciding unifiability of $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unification problems w.r.t. cycle-restricted $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-ontologies is PSpace-complete.*

To obtain this result, we combine the approach for unification in $\mathcal{EL}^{-\top}$ [8, 9] with the one for unification in $\mathcal{ELH}_{\mathcal{R}+}$ w.r.t. cycle-restricted ontologies [7, 12, 13], to devise a non-deterministic polynomial space unification algorithm for $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$ w.r.t. cycle-restricted ontologies. This algorithm follows the line of the one for $\mathcal{EL}^{-\top}$ in that it basically first generates $\mathcal{ELH}_{\mathcal{R}+}$-unifiers, which it then tries to pad with particles. Appropriate particles are found as solutions of certain linear language inclusions. However, due to the presence of GCIs and role axioms, quite a number of non-trivial changes and additions are required.

In particular, the solutions of the systems of linear language inclusions as constructed in [8, 9] cannot capture particles that are appropriate due to the presence of an ontology. For instance, in our example, *UpperBodyStructure* would be such a particle. To repair this problem, we first need to show that, in $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$, unifiability w.r.t. a cycle-restricted ontology can be characterized by the existence of a special type of unifiers. Afterwards, we exploit the properties of this kind of unifiers to define more sophisticated systems of language inclusions, which encode the semantics of GCIs and role axioms occurring in a background ontology. The solutions of such systems then yield also particles that are appropriate only due to the presence of this ontology.

Future Work

With SNOMED CT in mind, it would be interesting to see whether results on unification (with or without top) can be further extended to ontologies additionally containing so-called right-identity rules, i.e., role axioms of the form $r \circ s \sqsubseteq r$, since they are also needed to get rid of the SEP-triplet encoding mentioned above. However, this would require to extend the characterization of subsumption (between $\mathcal{EL}$ concepts w.r.t. a background ontology) to this setting, which is probably a non-trivial problem. From a theoretical point of view, the big open problem is whether one can dispense with the requirement that the ontology must be cycle-restricted. Even for pure $\mathcal{EL}$, decidability of unification w.r.t. unrestricted ontologies is an open problem.

From a practical point of view, the next step is to develop an algorithm that replaces non-deterministic guessing by a more intelligent search procedure. Since the unification problem is PSpace-complete, a polynomial-time translation of the whole problem into SAT is not possible (unless $\text{NP} = \text{PSpace}$). However, one could try to delegate the search for an $\mathcal{ELH}_{\mathcal{R}^+}$-unifier to a SAT solver, which interacts with a solver for the additional condition on such a unifier (existence of a certain type of solution for the corresponding system of linear language inclusions) in an SMT-like fashion [14].

The paper summarized by this abstract has been accepted for publication at IJCAR 2024 [15]. The extended version [16] of [15] provides detailed proof of our results.

References

[1] F. Baader, P. Narendran, Unification of concept terms in description logics, J. Symb. Comput. 31 (2001) 277–305.

[2] F. Baader, S. Borgwardt, B. Morawska, Constructing SNOMED CT Concepts via Disunification, LTCS-Report 17-07, Chair for Automata Theory, Institute for Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany, 2017. doi:10.25368/2022.237.

[3] F. Baader, B. Morawska, Unification in the description logic $\mathcal{EL}$, in: Rewriting Techniques and Applications, 20th International Conference, RTA 2009, Proceedings, volume 5595 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 350–364.

[4] F. Baader, B. Morawska, Unification in the description logic $\mathcal{EL}$, Log. Methods Comput. Sci. 6 (2010).

[5] F. Baader, B. Morawska, SAT encoding of unification in $\mathcal{EL}$, in: Logic for Programming, Artificial Intelligence, and Reasoning - 17th International Conference, LPAR-17, Proceedings, volume 6397 of *Lecture Notes in Computer Science*, Springer, 2010, pp. 97–111.

[6] F. Baader, J. Mendez, B. Morawska, UEL: unification solver for the description logic $\mathcal{EL}$ – system description, in: Automated Reasoning - 6th International Joint Conference, IJCAR 2012, Proceedings, volume 7364 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 45–51.

[7] F. Baader, S. Borgwardt, B. Morawska, Extending unification in $\mathcal{EL}$ towards general tboxes, in: Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012, AAAI Press, 2012.

[8] F. Baader, N. T. Binh, S. Borgwardt, B. Morawska, Unification in the description logic $\mathcal{EL}$ without the top concept, in: Proceedings of the 23rd International Conference on Automated Deduction (CADE 2011), volume 6803 of *Lecture Notes in Computer Science*, Springer-Verlag, 2011, pp. 70–84.

[9] F. Baader, T. B. Nguyen, S. Borgwardt, B. Morawska, Deciding unifiability and computing local unifiers in the description logic $\mathcal{EL}$ without top constructor, Notre Dame J. Formal Log. 57 (2016) 443–476.

[10] S. Schulz, M. Romacker, U. Hahn, Part-whole reasoning in medical ontologies revisited—introducing SEP triplets into classification-based description logics, in: AMIA 1998, American Medical Informatics Association Annual Symposium, AMIA, 1998.

[11] B. Suntisrivaraporn, F. Baader, S. Schulz, K. A. Spackman, Replacing SEP-triplets in SNOMED CT using tractable description logic operators, in: 11th Conference on Artificial Intelligence in Medicine, AIME 2007, Proceedings, volume 4594 of *Lecture Notes in Computer Science*, Springer, 2007, pp. 287–291.

[12] F. Baader, S. Borgwardt, B. Morawska, SAT encoding of unification in $\mathcal{ELH}_{\mathcal{R}^+}$ w.r.t. cycle-restricted ontologies, in: Automated Reasoning – 6th International Joint Conference, IJCAR 2012, Proceedings, volume 7364 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 30–44.

[13] F. Baader, S. Borgwardt, B. Morawska, A goal-oriented algorithm for unification in $\mathcal{ELH}_{\mathcal{R}^+}$ w.r.t. cycle-restricted ontologies, in: AI 2012: Advances in Artificial Intelligence – 25th Australasian Joint Conference, Proceedings, volume 7691 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 493–504.

[14] C. W. Barrett, R. Sebastiani, S. A. Seshia, C. Tinelli, Satisfiability modulo theories, in: A. Biere, M. Heule, H. van Maaren, T. Walsh (Eds.), Handbook of Satisfiability - Second Edition, volume 336 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2021, pp. 1267–1329. doi:10.3233/FAIA201017.

[15] F. Baader, O. Fernández Gil, Unification in the Description Logic $\mathcal{ELH}_{\mathcal{R}+}$ without the Top Concept modulo Cycle-Restricted Ontologies, in: Proc. of the 12th Int. Joint Conf. on Automated Reasoning, IJCAR 2024, 2024. To appear.

[16] F. Baader, O. Fernández Gil, Unification in the Description Logic $\mathcal{ELH}_{\mathcal{R}^+}$ without the Top Concept modulo Cycle-Restricted Ontologies (Extended Version), LTCS-Report 24-01, Chair for Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany, 2024. URL: https://lat.inf.tu-dresden.de/research/reports.html.

Unification in the Description Logic $\mathcal{ELH}_{\mathcal{R}^+}$ without the Top Concept modulo Cycle-Restricted Ontologies

Franz Baader and Oliver Fernández Gil

Institute of Theoretical Computer Science, TU Dresden, Dresden, Germany
`{franz.baader,oliver.fernandez}@tu-dresden.de`

Abstract. Unification has been introduced in Description Logic (DL) as a means to detect redundancies in ontologies. In particular, it was shown that testing unifiability in the DL $\mathcal{EL}$ is an NP-complete problem, and this result has been extended in several directions. Surprisingly, it turned out that the complexity increases to PSpace if one disallows the use of the top concept in concept descriptions. Motivated by features of the medical ontology SNOMED CT, we extend this result to a setting where the top concept is disallowed, but there is a background ontology consisting of restricted forms of concept and role inclusion axioms. We are able to show that the presence of such axioms does not increase the complexity of unification without top, i.e., testing for unifiability remains a PSpace-complete problem.

1 Introduction

Description Logics (DLs) [10] are a prominent family of logic-based knowledge representation languages, which offer their users a good compromise between expressiveness and complexity of reasoning, and constitute the formal and algorithmic foundation of the standard Web Ontology Language OWL 2.[1] The DL $\mathcal{EL}$, which provides the concept constructors conjunction ($\sqcap$), existential restriction ($\exists r.C$), and top concept ($\top$), is a rather inexpressive, but nevertheless very useful member of this family. On the one hand, the important reasoning problems, such as the subsumption and the equivalence problem, in $\mathcal{EL}$ and some of its extensions are decidable in polynomial time [22,8]. On the other hand, $\mathcal{EL}$ and its tractable extensions are frequently used to define biomedical ontologies, such as the large medical ontology SNOMED CT.[2] To illustrate the use of the top concept, whose absence plays an important rôle in this paper, consider the $\mathcal{EL}$ concept descriptions *Man* $\sqcap \exists child.\top$ and *Man* $\sqcap \exists child.Female$ of the concepts *Father* and *Father of a daughter*, respectively. In the former description, the top concept is used since no further properties of the child are to be required.

Unification in DLs has been introduced in [17] as a new inference service, motivated by the need for detecting redundancies in ontologies, in a setting where

---

different ontology engineers (OEs) constructing the ontology may model the same concepts on different levels of granularity. For example, assume that (using the style of SNOMED CT definitions) one OE models the concept of a *viral infection of the lung* as *ViralInfection* $\sqcap$ $\exists findingSite.LungStructure$ whereas another one models it as *LungInfection* $\sqcap \exists causativeAgent.Virus$. Here *ViralInfection* and *LungInfection* are used as atomic concepts without further defining them, i.e., the two OEs made different decisions when to stop the modelling process. The resulting concept descriptions are not equivalent, but they are nevertheless meant to represent the same concept. They can be made equivalent by treating the concept names *ViralInfection* and *LungInfection* as variables, and then substituting the first one by *Infection* $\sqcap \exists causativeAgent.Virus$ and the second one by *Infection* $\sqcap \exists findingSite.LungStructure$. In this case, we say that the descriptions are unifiable, and call the substitution that makes them equivalent a *unifier*. Intuitively, such a unifier proposes definitions for the concept names that are used as variables. In [7], unification and its extension to disunification are used to construct new medical concepts from SNOMED CT.

Unification in $\mathcal{EL}$ was first investigated in [14], where it was proved that deciding unifiability is an NP-complete problem. The NP upper bound was shown in that paper using a brute-force "guess and then test" NP algorithm. More practical algorithms for solving this problem and for computing unifiers were presented in [16] and [15], where the former describes a goal-oriented transformation-based algorithm and the latter is based on a translation to SAT. Implementations of these two algorithms are provided by the system UEL[3] [13], which is also available as a plug-in for the ontology editor Protégé. At the time these algorithms were developed, SNOMED CT was an $\mathcal{EL}$ ontology consisting of acyclic concept definitions. Since such definitions can be encoded into the unification problem (see Section 2.3 in [16]), algorithms for unification of $\mathcal{EL}$ concept descriptions (without background ontology) could be applied to SNOMED CT.

There was, however, one problem with employing these algorithms in the context of SNOMED CT: the top concept is not used in SNOMED CT, but the concepts generated by $\mathcal{EL}$ unification might contain $\top$, even if applied to concept descriptions not containing $\top$. Thus, the concept descriptions produced by the unifier are not necessarily in the style of SNOMED CT. For example, assume that we are looking for a unifier satisfying the two subsumption constraints[4] $\exists findingSite.LungStructure \sqsubseteq^? \exists findingSite.X$, $\exists findingSite.HeartStructure \sqsubseteq^? \exists findingSite.X$. It is easy to see that there is only one unifier of these two constraints, which replaces X with $\top$. Unification in $\mathcal{EL}^{-\top}$, i.e., the fragment of $\mathcal{EL}$ in which the top constructor is disallowed, was investigated in [1,18]. Surprisingly, it turned out that the absence of $\top$ makes unification considerably harder, both from a conceptual and a computational complexity point of view.

---

[3] https://sourceforge.net/projects/uel/

[4] Instead of equivalence constraints, as in our above example and in early work on unification in DLs, we consider here a set of subsumption constraints as unification problem. It is easy to see that these two kinds of unification problems can be reduced to each other [2].

In fact, the complexity of deciding unifiability increases from NP-complete for $\mathcal{EL}$ to PSpace-complete for $\mathcal{EL}^{-\top}$. The unification algorithm for $\mathcal{EL}^{-\top}$ introduced in [1,18] basically proceeds as follows. It first applies the unification algorithm for $\mathcal{EL}$ to compute so-called local unifiers. If none of them is an $\mathcal{EL}^{-\top}$-unifier, then it tries to pad the images of the variables by conjoining concept descriptions called particles. The task of finding appropriate particles is reduced to solving certain systems of linear language inclusions, which can be realized in PSpace using an automata-based approach.

The current version of SNOMED CT consists not only of acyclic concept definitions, but also contains more general concept inclusions (GCIs). In addition, properties of the part-of relation are no longer encoded using the so-called SEP-triplet encoding [27], but are directly expressed via role axioms [29], which can, for instance, be used to state that the part-of relation is transitive and that proper-part-of is a subrole of part-of. Decidability of unification in $\mathcal{EL}$ w.r.t. a background ontology consisting of GCIs is still an open problem. In [2], it is shown that the problem remains in NP if the ontology is cycle-restricted, which is a condition that the current version of SNOMED CT satisfies. Extensions of this result to the DL $\mathcal{ELH}_{\mathcal{R}+}$, which additionally allows for transitive roles and role inclusion axioms, were presented in [5] and [3], where the former introduces a SAT-based algorithm and the latter a transformation-based one. However, in all these algorithms, unifiers may introduce concept descriptions containing $\top$. In our example with the different finding site, however, the presence of the GCIs $LungStructure \sqsubseteq UpperBodyStructure$ and $HeartStructure \sqsubseteq UpperBodyStructure$ would yield a unifier not using $\top$, namely the one that replaces X with $UpperBodyStructure$.

The purpose of this paper is to combine the approach for unification in $\mathcal{EL}^{-\top}$ [1,18] with the one for unification in $\mathcal{ELH}_{\mathcal{R}+}$ w.r.t. cycle-restricted ontologies [2,5,3], to obtain a unification algorithm for $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$ w.r.t. cycle-restricted ontologies. This algorithm follows the line of the one for $\mathcal{EL}^{-\top}$ in that it basically first generates $\mathcal{ELH}_{\mathcal{R}+}$-unifiers, which it then tries to pad with particles. Appropriate particles are found as solutions of certain linear language inclusions. However, due to the presence of GCIs and role axioms, quite a number of non-trivial changes and additions are required. In particular, the solutions of the systems of linear language inclusions as constructed in [1,18] cannot capture particles that are appropriate due to the presence of an ontology. For instance, in our example, $UpperBodyStructure$ would be such a particle. To repair this problem, we first need to show that, in $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$, unifiability w.r.t. a cycle-restricted ontology can be characterized by the existence of a special type of unifiers. Afterwards, we exploit the properties of this kind of unifiers to define more sophisticated systems of language inclusions, which encode the semantics of GCIs and role axioms occurring in a background ontology. The solutions of such systems then yield also particles that are appropriate only due to the presence of this ontology.

While the unification problem investigated in this paper is motivated by an application in ontology engineering, it is also of interest for unification theory [19], which is concerned with unification-related properties of equational

theories. In fact, unification in DLs can be seen as a special case of unification modulo equational theories, where the respective equational theory axiomatizes equivalence in the DL under consideration. For $\mathcal{EL}$ and $\mathcal{ELH}_{\mathcal{R}^+}$, the corresponding equational theories can be found in [28]. The ones for the case without top can be obtained from them by removing the constant 1 from the signature, and all identities containing it from the axiomatization. The results in [1,18] and in the present paper show that the seemingly harmless removal of a constant from the equational theory may increase the complexity of the unification problem considerably. Considering unification w.r.t. a background ontology corresponds to adding a finite set of ground identities to the corresponding equational theory. For the word problem, it was shown that decidability is stable under adding finite sets of ground identities to theories such as commutativity or associativity-commutativity [25,20,11,24]. For unification, it was shown in [12] that adding finite sets of ground identities to the theory $ACUI$ of an associativity-commutativity-idempotent symbol with a unit leaves the unification problem decidable. The results in [2,5,3] can be seen as such transfer results, but they require a restriction on the ground identities corresponding to cycle-restrictedness.

Due to space constraints, we cannot give detailed proof of our results here. They can be found in [9] (see also the appendix).

2   Subsumption and Unification in $\mathcal{ELH}_{\mathcal{R}^+}$ and $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$

First, we briefly introduce syntax and semantics of the DLs investigated in this paper. Then, we recall a useful characterization of subsumption for these logics, and finally define the unification problem.

2.1   The DLs $\mathcal{ELH}_{\mathcal{R}^+}$ and $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$

Starting with countably infinite sets $\mathsf{N_C}$ and $\mathsf{N_R}$ of concept names and role names, $\mathcal{ELH}_{\mathcal{R}^+}$-*concept descriptions* (for short, *concepts*) are built using the concept constructors *conjunction* ($\sqcap$), *existential restriction* ($\exists r.C$), and *top* ($\top$). When building $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-concepts, the constructor $\top$ is not available. An $\mathcal{ELH}_{\mathcal{R}^+}$-ontology $\mathcal{O}$ is a finite set of *general concept inclusions (GCIs) $C \sqsubseteq D$, role hierarchy axioms $r \sqsubseteq s$*, and *transitivity axioms $r \circ r \sqsubseteq r$*, where C, D are $\mathcal{ELH}_{\mathcal{R}^+}$-concepts and r, s are role names. In an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-ontology, the concepts occurring in GCIs must be $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-concepts.

The following two notions will play an important rôle in our unification algorithm. An *atom* is either a concept name or an existential restriction, and a *particle* is an atom of the form $\exists r_1.\exists r_2. \cdots \exists r_n.A$ for a concept name A, which we write as $\exists w.A$, where $w = r_1 \ldots r_n$ is viewed as a word over the alphabet $\mathsf{N_R}$. Every $\mathcal{ELH}_{\mathcal{R}^+}$-concept C is a conjunction of atoms, where the empty conjunction represents $\top$. These atoms are called the *top-level atoms* of C. The set $Ats(C)$ consists of all atoms (not just top-level ones) occurring

in C, and $Ats(\mathcal{O})$ for an ontology $\mathcal{O}$ consists of the atoms of all concepts occurring in $\mathcal{O}$. The set of particles of an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-concept is defined inductively: $Part(A) := \{A\}$ for each concept name A, $Part(\exists r.C) := \{\exists r.P \mid P \in Part(C)\}$, and $Part(C \sqcap D) := Part(C) \cup Part(D)$. For example, if $C = \exists r.(\exists s.A \sqcap \exists r.B)$, then $Part(C) = \{\exists rs.A, \exists rr.B\}$ and $Ats(C) = \{C, \exists s.A, \exists r.B, A, B\}$, where C is the only top-level atom.

The *semantics* of $\mathcal{ELH}_{\mathcal{R}+}$-concepts and ontologies is defined using the notion of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, which has a set $\Delta^{\mathcal{I}} \neq \emptyset$ as interpretation domain, and assigns a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ to each concept name A and a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ to each role name r. The interpretation function $\cdot^{\mathcal{I}}$ is extended to $\mathcal{ELH}_{\mathcal{R}+}$-concepts as usual: $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}$, $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$, and $(\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \exists e.((d,e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}})\}$. The interpretation $\mathcal{I}$ is a *model* of the $\mathcal{ELH}_{\mathcal{R}+}$-ontology $\mathcal{O}$ if $C \sqsubseteq D \in \mathcal{O}$ implies $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, $r \sqsubseteq s \in \mathcal{O}$ implies $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$, and $r \circ r \sqsubseteq r \in \mathcal{O}$ implies that $r^{\mathcal{I}}$ is transitive.

2.2 Subsumption in $\mathcal{ELH}_{\mathcal{R}+}$ and $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$

Given an $\mathcal{ELH}_{\mathcal{R}+}$-ontology $\mathcal{O}$ and $\mathcal{ELH}_{\mathcal{R}+}$-concepts C, D, we say that C is *subsumed* by D w.r.t. $\mathcal{O}$ (written $C \sqsubseteq_{\mathcal{O}} D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{O}$. They are *equivalent* w.r.t. $\mathcal{O}$ (written $C \equiv_{\mathcal{O}} D$) if $C \sqsubseteq_{\mathcal{O}} D$ and $D \sqsubseteq_{\mathcal{O}} C$.

Subsumption (and thus also equivalence) between $\mathcal{ELH}_{\mathcal{R}+}$-concepts w.r.t. arbitrary $\mathcal{ELH}_{\mathcal{R}+}$-ontologies can be decided in polynomial time [8]. In the context of unification, a recursive characterization of subsumption turns out to be useful, which for $\mathcal{ELH}_{\mathcal{R}+}$ was first given in [5], and later reformulated in [3]. In this paper we use the one given in [3], but before we can formulate this characterization, we must introduce the *role hierarchy* induced by an $\mathcal{ELH}_{\mathcal{R}+}$-ontology $\mathcal{O}$: given role names r, s, we say that r is a *subrole* of s (written $r \trianglelefteq_{\mathcal{O}} s$) if $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathcal{O}$. It is easy to see that the relation $\trianglelefteq_{\mathcal{O}}$ is the reflexive-transitive closure of the explicitly stated subrole relationships $\{(r,s) \mid r \sqsubseteq s \in \mathcal{O}\}$. We call a role name r *transitive* if $r \circ r \sqsubseteq r \in \mathcal{O}$.

The characterization of subsumption in [3] uses the notion of *structural subsumption*: given atoms C, D, we say that C *is structurally subsumed by* D w.r.t. an $\mathcal{ELH}_{\mathcal{R}+}$-ontology $\mathcal{O}$ (written $C \sqsubseteq_{\mathcal{O}}^{s} D$) if one of the following cases applies:

1. $C = D$ is a concept name.
2. $C = \exists r.C'$, $D = \exists s.D'$, $r \trianglelefteq_{\mathcal{O}} s$, and $C' \sqsubseteq_{\mathcal{O}} D'$.
3. $C = \exists r.C'$, $D = \exists s.D'$, and $C' \sqsubseteq_{\mathcal{O}} \exists t.D'$ for some transitive role name t satisfying $r \trianglelefteq_{\mathcal{O}} t \trianglelefteq_{\mathcal{O}} s$.

Lemma 1 ([3]). *Let $\mathcal{O}$ be an $\mathcal{ELH}_{\mathcal{R}+}$-ontology and $C_1, \ldots, C_n, D_1, \ldots, D_m$ atoms. Then, $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq_{\mathcal{O}} D_1 \sqcap \cdots \sqcap D_m$ iff for every $j \in \{1, \ldots, m\}$:*

1. *there is an index $i \in \{1, \ldots, n\}$ such that $C_i \sqsubseteq_{\mathcal{O}}^{s} D_j$, or*
2. *there are atoms $At_1, \ldots, At_k, At'$ of $\mathcal{O}$ $(k \geq 0)$ such that:*
   (a) $At_1 \sqcap \cdots \sqcap At_k \sqsubseteq_{\mathcal{O}} At'$,
   (b) for every $\ell \in \{1, \ldots, k\}$ there exists $i \in \{1, \ldots, n\}$ with $C_i \sqsubseteq_{\mathcal{O}}^{s} At_\ell$, and
   (c) $At' \sqsubseteq_{\mathcal{O}}^{s} D_j$.

If $\mathcal{O}$ is empty, then the second case in the definition of structural subsumption can be modified to require that $r = s$ and $C' \sqsubseteq_\emptyset D'$, whereas the third case in the same definition as well as the second case in Lemma 1 can be removed. This then yields the characterization of subsumption in $\mathcal{EL}$ of [16]. Since $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$ is a fragment of $\mathcal{ELH}_{\mathcal{R}+}$, this characterization also applies to subsumption between $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-concepts w.r.t. $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-ontologies. However, in this setting, the case $k = 0$ in 2. cannot occur. This is a direct consequence of the following result.

Lemma 2. *If $\mathcal{O}$ is an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-ontology and At an atom of $\mathcal{O}$, then $\top \not\sqsubseteq_\mathcal{O} At$.*

2.3   Unification in $\mathcal{ELH}_{\mathcal{R}+}$ and $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$

When defining unification, we assume that the set of concept names is partitioned into a set $\mathsf{N_C}$ of concept constants and a set $\mathsf{N_V}$ of concept variables. Given a DL $\mathcal{L} \in \{\mathcal{ELH}_{\mathcal{R}+}, \mathcal{ELH}_{\mathcal{R}+}^{-\top}\}$, an $\mathcal{L}$-*substitution* σ is a mapping from a finite subset of $\mathsf{N_V}$ to the set of $\mathcal{L}$-concepts. The application of σ to an arbitrary $\mathcal{L}$-concept is defined inductively in the usual way. A concept (ontology) is *ground* if it does not contain variables. A substitution σ is ground if $\sigma(X)$ is ground for all variables X that have an image under σ.

Definition 1. *Let $\mathcal{O}$ be a ground ontology. An $\mathcal{L}$-unification problem w.r.t. $\mathcal{O}$ is of the form $\Gamma = \{C_1 \sqsubseteq^? D_1, \ldots, C_n \sqsubseteq^? D_n\}$, where $C_1, D_1, \ldots, C_n, D_n$ are $\mathcal{L}$-concepts. An $\mathcal{L}$-substitution σ is an $\mathcal{L}$-unifier of Γ w.r.t. $\mathcal{O}$ if $\sigma(C_i) \sqsubseteq_\mathcal{O} \sigma(D_i)$ for all $i \in \{1, \ldots, n\}$. The unification problem Γ is called $\mathcal{L}$-unifiable w.r.t. $\mathcal{O}$ if it has an $\mathcal{L}$-unifier w.r.t. $\mathcal{O}$.*

The following example illustrates that unifiability of a given unification problem may depend on the considered DL $\mathcal{L}$ and on the presence of a non-empty ontology.

Example 1. Let $\mathcal{O} = \emptyset$ and consider the following unification problem:

$$\Gamma_1 := \{\exists r.A \sqsubseteq^? X, \quad \exists u.B \sqsubseteq^? Y, \quad \exists s.X \sqcap A \sqsubseteq^? Y\}.$$

Viewed as an $\mathcal{ELH}_{\mathcal{R}+}$-unification problem, it has the unifier σ with $\sigma(X) = \sigma(Y) = \top$. However, Γ_1 does not have an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier w.r.t. $\mathcal{O} = \emptyset$. To see this, suppose that δ is such a unifier. Using Lemma 1 for the special case of an empty ontology, we can deduce from $\exists u.B \sqsubseteq_\emptyset \delta(Y)$ that every top-level atom of $\delta(Y)$ is an existential restriction for the role u. However, we can also deduce from $\exists s.\delta(X) \sqcap A \sqsubseteq_\emptyset \delta(Y)$ that every top-level atom of $\delta(Y)$ is either A or an existential restriction for the role s. Since not both is possible, $\delta(Y)$ cannot have any top-level atoms, and thus must be $\top$, contradicting our assumption that δ is an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier. If we define $\mathcal{O}' := \{B \sqsubseteq \exists r.A, \ u \sqsubseteq s\}$, then the $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifiability status of Γ_1 changes to unifiable since δ with $\delta(X) = \exists r.A$ and $\delta(Y) = \exists s.\exists r.A$ is an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier of Γ_1 w.r.t. $\mathcal{O}'$.

In the next section we will show how to decide unifiability of an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unification problem w.r.t. a cycle-restricted $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-ontology.

Definition 2. *An $\mathcal{ELH}_{\mathcal{R}+}$-ontology $\mathcal{O}$ is called* cycle-restricted *if there is no sequence of $n > 0$ role names $r_1, \ldots, r_n \in \mathsf{N_R}$ and $\mathcal{ELH}_{\mathcal{R}+}$-concept C such that $C \sqsubseteq_{\mathcal{O}} \exists r_1.\exists r_2.\cdots\exists r_n.C$.*

As stated in [5] (and proved in [6]), one can test in polynomial time whether a given $\mathcal{ELH}_{\mathcal{R}+}$-ontology is cycle-restricted or not.

According to [5,18], we can without loss of generality assume that the given ontology and the unification problem are *flat*. An $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-atom is flat if it is a concept name or of the form $\exists r.A$ for a concept name A. A GCI $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq D$ or subsumption constraint $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? D$ is flat if $C_1, \ldots, C_n$ and D are flat $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-atoms. Finally, an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-ontology or $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unification problem is flat if all it elements are flat.

The following result for flat, cycle-restricted $\mathcal{ELH}_{\mathcal{R}+}$-ontologies will turn out to be quite useful in the next section. It basically follows from the proof of Lemma 8 in [4].

Lemma 3. *Let $\mathcal{O}$ be a flat, cycle-restricted $\mathcal{ELH}_{\mathcal{R}+}$-ontology, $A \in \mathsf{N_C}$ and $\exists r.C$ an $\mathcal{ELH}_{\mathcal{R}+}$-atom. Then, $A \sqsubseteq_{\mathcal{O}} \exists r.C$ iff there exists $\exists u.B \in Ats(\mathcal{O})$ such that $B \sqsubseteq_{\mathcal{O}} C$, and*

 - *$A \sqsubseteq_{\mathcal{O}} \exists u.B$ and $u \trianglelefteq_{\mathcal{O}} r$, or*
 - *$A \sqsubseteq_{\mathcal{O}} \exists t.B$ for a transitive role t with $u \trianglelefteq_{\mathcal{O}} t \trianglelefteq_{\mathcal{O}} r$.*

3   The Unification Algorithm for $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$

In the following, we assume that $\mathcal{O}$ is a flat and cycle-restricted $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-ontology and Γ is a flat $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unification problem. We introduce an algorithm that can test whether Γ has an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier and needs only polynomial space for this task. This algorithm follows the approach developed in [18] for unification in $\mathcal{EL}^{-\top}$, but must take the ontology into account, which means that it must deal with a considerably more complex characterization of subsumption (see Lemma 1 and our remarks on how the characterization can be simplified if $\mathcal{O} = \emptyset$).

Before presenting our new approach, we briefly sketch the one employed in [18]. The original NP procedure for unification in $\mathcal{EL}$ [16] is based on the (non-trivial) observation that an $\mathcal{EL}$-unification problem Γ has a unifier iff it has a *local unifier*, i.e., one that is built using only atoms occurring in the unification problem. The procedure guesses an appropriate representation of a local substitution, and then checks by $\mathcal{EL}$ reasoning whether it really is a unifier. Basically, to guess a local substitution σ, one must guess for every variable X and non-variable atom C of Γ whether $\sigma(X) \sqsubseteq_{\emptyset} \sigma(C)$ is supposed to hold. A *subsumption mapping* τ describing a local unifier σ more generally guesses for every pair C, D of atoms whether $\sigma(C) \sqsubseteq_{\emptyset} \sigma(D)$ is supposed to hold. The restrictions imposed on such subsumption mappings ensure that the local substitution induced by such a mapping is indeed an $\mathcal{EL}$-unifier of Γ [18], i.e., the subsequent $\mathcal{EL}$ reasoning testing this can be dispensed with. The local unifier obtained from

a subsumption mapping τ need not be an $\mathcal{EL}^{-\top}$-unifier. To test for the existence of an $\mathcal{EL}^{-\top}$-unifier related to τ, the subsumption mapping τ together with the original unification problem Γ is then used to construct a new unification problem $\Delta_{\Gamma,\tau}$, in which only variables can occur on the right-hand side of subsumption constraints. Existence of an $\mathcal{EL}^{-\top}$-unifier of $\Delta_{\Gamma,\tau}$ that is compatible with τ is then reduced in [18] to the existence of an admissible solution of a corresponding set $\mathfrak{I}_{\Gamma,\tau}$ of linear language inclusions. The latter problem can in turn be reduced in polynomial time to checking emptiness of alternating finite automata with ε-transitions [18], which is a PSpace-complete problem [23].

In this section we show how this approach can be extended from $\mathcal{EL}^{-\top}$ to $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$ w.r.t. cycle-restricted ontologies. We start by introducing subsumption mappings and the induced unification problems of the form $\Delta_{\Gamma,\tau}$.

3.1   The Subsumption Mapping

Let $Ats(\Gamma, \mathcal{O})$ be the set of atoms occurring in Γ or $\mathcal{O}$. Due to the third case in the definition of structural subsumption, we also need to consider certain atoms that are not explicitly present in the input:

$$Ats_{tr}(\Gamma, \mathcal{O}) := Ats(\Gamma, \mathcal{O}) \cup \{\exists t.C \mid \exists s.C \in Ats(\Gamma, \mathcal{O}),\ t \trianglelefteq_{\mathcal{O}} s,\ t \text{ is transitive}\}.$$

A *non-variable atom* is an atom in $Ats_{tr}(\Gamma, \mathcal{O})$ that is not a variable. We denote the set of all such atoms as $At_{nv}(\Gamma, \mathcal{O})$. A mapping of the form $\tau : Ats_{tr}(\Gamma, \mathcal{O}) \times Ats_{tr}(\Gamma, \mathcal{O}) \to \{0, 1\}$ induces an assignment S^{τ} that maps variables in Γ to sets of non-variable atoms in $Ats_{tr}(\Gamma, \mathcal{O})$:

$$S^{\tau}(X) := \{D \in At_{nv}(\Gamma, \mathcal{O}) \mid \tau(X, D) = 1\}.$$

This assignment induces the relation $>_{S^{\tau}} := \{(X, Y) \in Vars(\Gamma) \times Vars(\Gamma) \mid Y \text{ occurs in an atom of } S^{\tau}(X)\}$. We say that S^{τ} is *acyclic* if the transitive closure of $>_{S^{\tau}}$ is irreflexive, and thus a strict partial order, which we denote as $>_{\tau}$. If S^{τ} is acyclic, then it induces a substitution σ_{τ}, defined by induction on $>_{\tau}$:

- If X is minimal w.r.t. $>_{\tau}$, then $\sigma_{\tau}(X) := \prod_{D \in S^{\tau}(X)} D$.
- Otherwise, assuming that $\sigma_{\tau}(Y)$ has already been defined for all Y such that $X >_{\tau} Y$, one defines $\sigma_{\tau}(X) := \prod_{D \in S^{\tau}(X)} \sigma_{\tau}(D)$.

The conditions imposed on a subsumption mapping τ ensure that the induced substitution σ_{τ} is an $\mathcal{ELH}_{\mathcal{R}^+}$-unifier of Γ. In order to simplify the definition of these conditions, we introduce the following notation (for atoms $\exists r.C, \exists s.D$): $\mathcal{S}(\exists r.C, \exists s.D) := \{D \mid \text{if } r \trianglelefteq_{\mathcal{O}} s\} \cup \{\exists t.D \mid r \trianglelefteq_{\mathcal{O}} t \trianglelefteq_{\mathcal{O}} s,\ t \text{ transitive}\}$. Basically, this set collects all concepts F such that $C \sqsubseteq_{\mathcal{O}} F$ implies $\exists r.C \sqsubseteq_{\mathcal{O}}^{s} \exists s.D$ (see the second and third case in the definition of $\sqsubseteq_{\mathcal{O}}^{s}$).

Definition 3. *The mapping* $\tau : Ats_{tr}(\Gamma, \mathcal{O}) \times Ats_{tr}(\Gamma, \mathcal{O}) \to \{0, 1\}$ *is called a subsumption mapping for* Γ *w.r.t.* $\mathcal{O}$ *if it satisfies the following conditions:*

1. *It respects the properties of subsumption w.r.t.* $\mathcal{O}$:

(a) $\tau(D, D) = 1$, for each $D \in Ats_{tr}(\Gamma, \mathcal{O})$.

(b) For all $D_1, D_2, D_3 \in Ats_{tr}(\Gamma, \mathcal{O})$, if $\tau(D_1, D_2) = \tau(D_2, D_3) = 1$ then $\tau(D_1, D_3) = 1$.

(c) $\tau(C, D) = 1$ iff $C \sqsubseteq_{\mathcal{O}} D$, for all ground atoms $C, D \in Ats_{tr}(\Gamma, \mathcal{O})$.

(d) For each concept constant $A \in Ats(\Gamma, \mathcal{O})$, role name r, and variable X with $\exists r.X \in Ats_{tr}(\Gamma)$:

  i. $\tau(A, \exists r.X) = 1$ iff[5] there is an atom $\exists u.B$ of $\mathcal{O}$ such that $\tau(B, X) = 1$, and
    – $A \sqsubseteq_{\mathcal{O}} \exists u.B$ and $u \trianglelefteq_{\mathcal{O}} r$, or
    – $A \sqsubseteq_{\mathcal{O}} \exists t.B$ for a transitive role t with $u \trianglelefteq_{\mathcal{O}} t \trianglelefteq_{\mathcal{O}} r$.

  ii. $\tau(\exists r.X, A) = 1$ iff
    – there are atoms $\exists r_1.A_1, \ldots, \exists r_k.A_k$ of $\mathcal{O}$ $(k \geq 0)$ and atoms $F_\ell \in \mathcal{S}(\exists r.X, \exists r_\ell.A_\ell)$ $(1 \leq \ell \leq k)$ such that:
      $\tau(X, F_\ell) = 1$ $(1 \leq \ell \leq k)$ and $\exists r_1.A_1 \sqcap \cdots \sqcap \exists r_k.A_k \sqsubseteq_{\mathcal{O}} A$.

(e) For all role names $r, s \in \mathsf{N_R}$, variables X, and atoms $\exists r.C, \exists s.D \in Ats_{tr}(\Gamma)$ with $C = X$ or $D = X$: $\tau(\exists r.C, \exists s.D) = 1$ iff
  – there exists $F \in \mathcal{S}(\exists r.C, \exists s.D)$ such that $\tau(C, F) = 1$, or
  – there are atoms $\exists r_1.A_1, \ldots, \exists r_k.A_k, \exists u.B$ of $\mathcal{O}$ $(k \geq 0)$, atoms $F_\ell \in \mathcal{S}(\exists r.C, \exists r_\ell.A_\ell)$ $(1 \leq \ell \leq k)$, and an atom $F \in \mathcal{S}(\exists u.B, \exists s.D)$, such that: $\tau(C, F_\ell) = 1$ $(1 \leq \ell \leq k)$, $\exists r_1.A_1 \sqcap \cdots \sqcap \exists r_k.A_k \sqsubseteq_{\mathcal{O}} \exists u.B$, $\tau(B, F) = 1$.

2. The assignment S^τ is acyclic. Note that this means that τ induces the $\mathcal{ELH}_{\mathcal{R}+}$-substitution σ^τ.

3. The substitution σ^τ is an $\mathcal{ELH}_{\mathcal{R}+}$-unifier of Γ w.r.t. $\mathcal{O}$. In combination with the conditions already introduced, this is expressed by the following conditions for each subsumption constraint $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? D \in \Gamma$:

  (a) If D is a non-variable atom, then either $\tau(C_i, D) = 1$ for some $i \in \{1, \ldots, n\}$, or there are atoms $At_1, \ldots, At_k, At'$ of $\mathcal{O}$ $(k \geq 0)$ such that:
    – $At_1 \sqcap \cdots \sqcap At_k \sqsubseteq_{\mathcal{O}} At'$,
    – for each $\ell \in \{1, \ldots, k\}$ there is $i \in \{1, \ldots, n\}$ s.t. $\tau(C_i, At_\ell) = 1$, and
    – $\tau(At', D) = 1$.

  (b) If D is a variable and $\tau(D, C) = 1$ for a non-variable atom $C \in At_{nv}(\Gamma, \mathcal{O})$, then $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? C$ must satisfy the previous case.

By using the close relationship between this definition and the characterization of subsumption in Lemma 1, one can show that Γ has an $\mathcal{ELH}_{\mathcal{R}+}$-unifier w.r.t. $\mathcal{O}$ iff there is a subsumption mapping for Γ w.r.t. $\mathcal{O}$. In the proof of the if-direction, one shows that the substitution induced by the subsumption mapping is indeed a unifier. For the other direction, one takes a unifier σ and shows that the mapping τ satisfying $\tau(C, D) = 1$ iff $\sigma(C) \sqsubseteq_{\mathcal{O}} \sigma(D)$ is a subsumption mapping for Γ w.r.t. $\mathcal{O}$.

However, using subsumption mappings to characterize unifiability in $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$ requires more effort. Together with the unification problem Γ, a subsumption mapping τ yields a simpler unification problem $\Delta_{\Gamma, \tau} := \Delta_\Gamma \cup \Delta_\tau$, where

$$\Delta_\Gamma := \{C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? X \in \Gamma \mid X \in \mathsf{N_V}\} \quad \text{and} \quad \Delta_\tau := \{C \sqsubseteq^? X \mid \tau(C, X) = 1\}.$$

---

[5] This condition is justified by Lemma 3.

In addition, any substitution σ induces an assignment S^σ of the form:

$$S^\sigma(X) := \{D \in At_{nv}(\Gamma, \mathcal{O}) \mid \sigma(X) \sqsubseteq_{\mathcal{O}} \sigma(D)\}.$$

We write $S^\tau \leq S^\sigma$ if $S^\tau(X) \subseteq S^\sigma(X)$ holds for all variables X. In this case we say that σ is *compatible* with τ.

The following result gives a characterization of the existence of an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier w.r.t. an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-ontology.

Proposition 1. *Let $\mathcal{O}$ be a flat and cycle-restricted $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-ontology and Γ a flat $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unification problem. Then, Γ has an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier w.r.t. $\mathcal{O}$ iff there exists a subsumption mapping τ for Γ w.r.t. $\mathcal{O}$ such that $\Delta_{\Gamma,\tau}$ has an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier γ w.r.t. $\mathcal{O}$ that is compatible with τ.*

Example 2. Let $\mathcal{O} = \emptyset$ and consider the following unification problem:

$$\Gamma_2 := \{\exists r.B \sqsubseteq^? \exists r.Y, \quad \exists s.X \sqcap \exists r.A \sqsubseteq^? Y\}.$$

Due to Condition 3 in Definition 3 and the fact that $\mathcal{O}$ is empty, any subsumption mapping τ must satisfy $\tau(\exists r.B, \exists r.Y) = 1$. Condition 1e then implies that $\tau(B, Y) = 1$ must hold as well. Regarding the second subsumption constraint in Γ_2, Condition 3b does not apply as long as there is no non-variable atom C with $\tau(Y, C) = 1$. We can conclude that, for any subsumption mapping τ, the set $\Delta_{\Gamma_2,\tau}$ contains at least the subsumption constraints $B \sqsubseteq^? Y$ and $\exists s.X \sqcap \exists r.A \sqsubseteq^? Y$. Using an argument similar to the one employed in Example 1, one can show that such a set $\Delta_{\Gamma_2,\tau}$ cannot have an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier w.r.t. $\emptyset$.

It is easy to see that there also is a subsumption mapping τ that has only these two constraints in $\Delta_{\Gamma_2,\tau}$ since the only other mandatory values 1 are the ones required by 1a. For the ontology $\mathcal{O}'' = \{B \sqsubseteq \exists r.A\}$, the set $\Delta_{\Gamma_2,\tau}$ then has an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier w.r.t. $\mathcal{O}''$, which maps Y to $\exists r.A$. This unifier is compatible with τ since the subsumption mapping τ that yields value 1 only if required satisfies $S^\tau(X) = S^\tau(Y) = \emptyset$. Thus, by Proposition 1, Γ_2 has an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier w.r.t. $\mathcal{O}''$. Note that this unifier is not σ_τ since σ_τ in this case assigns $\top$ to X and Y.

3.2   Translation into Language Inclusions

Linear language inclusions are a special case of the linear language equations considered in [17] in the context of unification in the DL $\mathcal{FL}_0$. In contrast to the general case, where solvability is an ExpTime-complete problem [17], the linear language inclusions introduced in [18] in the context of unification in $\mathcal{EL}^{-\top}$ have a PSpace-complete solvability problem [18].

Definition 4. *Let $X_1, \ldots, X_n$ be a finite set of* indeterminates. *A linear language inclusion over this set of indeterminates and the alphabet $\mathsf{N_R}$ is an expression of the form*

$$X_i \subseteq L_0 \cup L_1 X_1 \cup \cdots \cup L_n X_n,$$

where $i \in \{1, \ldots, n\}$ and each $L_j \subseteq \{\varepsilon\} \cup \mathsf{N_R}$ $(0 \le j \le n)$. As usual, the symbol ε denotes the empty word. A solution θ of such an inclusion assigns sets of words $\theta(X_i) \subseteq \mathsf{N_R}^*$ to each indeterminate X_i such that $\theta(X_i) \subseteq L_0 \cup L_1 \cdot \theta(X_1) \cup \cdots \cup L_n \cdot \theta(X_n)$, where "$\cdot$" denotes concatenation of languages. The solution θ is finite *if $\theta(X_i)$ is a finite set for all $i \in \{1, \ldots, n\}$.*

Checking whether $\Delta_{\Gamma,\tau}$ has an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier w.r.t. $\mathcal{O}$ that is compatible with a given subsumption mapping τ can be reduced to solving a system $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ of such linear language inclusion. The basic idea is that, for each concept variable X and concept constant A, we introduce an indeterminate X_A. Intuitively, the system $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ is constructed such that the following holds:

- if γ is an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier of $\Delta_{\Gamma,\tau}$ compatible with τ, then there is an assignment θ_γ satisfying $\theta_\gamma(X_A) = \{w \mid \exists w.A \in Part(\gamma(X))\}$ that is a finite solution of the system $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$.

Since γ is an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier, of which we can assume without loss of generality that it is ground [19], the solution θ_γ satisfies an additional property: for every variable X there is a concept constant A such that $\theta_\gamma(X_A) \neq \emptyset$. We call a solution of $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ satisfying this property *admissible*. Conversely, finite, admissible solutions of $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ yield an appropriate unifier of $\Delta_{\Gamma,\tau}$:

- if $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ has a finite, admissible solution, then it has such a solution θ that yields an $\mathcal{ELH}_{\mathcal{R}+}^{-\top}$-unifier γ_θ of $\Delta_{\Gamma,\tau}$ that is compatible with τ. This unifier is defined similarly to σ_τ, but using particles provided by θ for padding:
  - if X is minimal w.r.t. $>_\tau$, then
    $$\gamma_\theta(X) := \bigsqcap_{D \in S^\tau(X)} D \sqcap \bigsqcap_{A \in \mathsf{N_C}} \bigsqcap_{w \in \theta(X_A)} \exists w.A,$$
  - if $\gamma_\theta(Y)$ has already been defined for all Y such that $X >_\tau Y$, then
    $$\gamma_\theta(X) := \bigsqcap_{D \in S^\tau(X)} \gamma_\theta(D) \sqcap \bigsqcap_{A \in \mathsf{N_C}} \bigsqcap_{w \in \theta(X_A)} \exists w.A.$$

Basically, to define the linear language inclusions in $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$, we consider the following situation: given a particle $\exists w.A \in Part(\gamma(X))$ and a constraint $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? X \in \Delta_{\Gamma,\tau}$, we know (by Lemma 2 in [18]) that $\gamma(C_1) \sqcap \cdots \sqcap \gamma(C_n) \sqsubseteq_\mathcal{O} \exists w.A$ holds. Hence, the idea is to encode, within the inclusions in $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$, whether a conjunction of atoms and a particle satisfy the characterization of subsumption in Lemma 1.

For the case of an empty ontology, the construction of the system $\mathfrak{I}_{\Gamma,\tau}^{\emptyset}$ is relatively straightforward since the characterization of subsumption is quite simple in this case. As described in [18], for each concept constant $A \in \mathsf{N_C}$ and each subsumption constraint $\mathfrak{s} = C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? X$ in $\Delta_{\Gamma,\tau}$, a linear inclusion $\mathsf{i}_A(\mathfrak{s})$ of the following form is added to $\mathfrak{I}_{\Gamma,\tau}^{\emptyset}$:

$$X_A \subseteq f_A(C_1) \cup \cdots \cup f_A(C_n), \quad \text{where } f_A(C) := \begin{cases} \{r\}f_A(C') & \text{if } C = \exists r.C', \\ Y_A & \text{if } C = Y \in \mathsf{N_V}, \\ \{\varepsilon\} & \text{if } C = A, \\ \emptyset & \text{if } C \in \mathsf{N_C} \setminus \{A\}. \end{cases}$$

Example 3. Consider the system $\Delta_{\Gamma_2,\tau} = \{B \sqsubseteq^? Y, \exists s.X \sqcap \exists r.A \sqsubseteq^? Y, \dots\}$ from Example 2. The first subsumption constraint yields the language inclusions $Y_A \subseteq \emptyset$ and $Y_B \subseteq \{\varepsilon\}$, and the second yields $Y_A \subseteq \{s\}X_A \cup \{r\}\{\varepsilon\}$ and $Y_B \subseteq \{s\}X_B \cup \{r\}\emptyset$. There are no language inclusions constraining X_A or X_B. Any solution θ of $\mathfrak{I}^{\emptyset}_{\Gamma_2,\tau}$ thus must satisfy $\theta(Y_A) = \emptyset$. If θ is admissible, then $\theta(Y_B)$ must be non-empty. The first inclusion for Y_B says that $\theta(Y_B)$ consists of the empty word, whereas the second says that every element of $\theta(Y_B)$ must start with the letter s. Thus, $\mathfrak{I}^{\emptyset}_{\Gamma_2,\tau}$ cannot have an admissible solution.

To take a non-empty ontology into account, the right-hand sides of the language inclusions must be extended. Our new translation yields linear language inclusions $\mathfrak{i}^*_A(\mathfrak{s})$ of the form

$$X_A \subseteq f^*_A(C_1) \cup \cdots \cup f^*_A(C_n) \cup \mathcal{U}_A(\mathfrak{s}), \tag{1}$$

where $f^*_A(C)$ differs from $f_A(C)$ in the way existential restrictions are treated: $f^*_A(\exists r.C') := L_r f_A(C')$ where $L_r := \{s \in \mathsf{N_R} \mid r \trianglelefteq_{\mathcal{O}} s\}$.

This modification of f_A to f^*_A takes care of the role hierarchy.

Example 4. For instance, if in the system of Example 3 we replace $B \sqsubseteq^? Y$ with $\exists u.X \sqsubseteq^? Y$, then the language inclusions corresponding to this constraint are $Y_A \subseteq \{u\}X_A$ and $Y_B \subseteq \{u\}X_B$. The new system again does not have an admissible solution. However, if we consider an ontology $\mathcal{O}$ containing $u \sqsubseteq s$, then the new translation yields the language inclusions $Y_A \subseteq \{u,s\}X_A$ and $Y_B \subseteq \{u,s\}X_B$ for this constraint. Consequently, the new system of language inclusions has a finite, admissible solution, which reflects the fact that the system of subsumption constraints has an $\mathcal{ELH}^{-\top}_{\mathcal{R}^+}$-unifier w.r.t. $\mathcal{O}$.

The GCIs and transitivity axioms of the ontology are taken care of by the additional term $\mathcal{U}_A(\mathfrak{s})$ in (1). This term uses additional types of indeterminates whose meaning is encoded using additional language inclusions. Indeterminates of the form $Z_{B \to A}$, where A, B are concept constants occurring in Γ or $\mathcal{O}$, are supposed to represent languages containing only words w such that $B \sqsubseteq_{\mathcal{O}} \exists w.A$. This intuition is formalized by the set of linear inclusions $\mathfrak{I}_{\mathcal{O}}$, which consists of one language inclusion for each indeterminate $Z_{B \to A}$ having the following form:

$$Z_{B \to A} \subseteq L \cup \bigcup_{(r,B') \in I(B)} \{r\}Z_{B' \to A}, \tag{2}$$

where $I(B) := \{(r, B') \in \mathsf{N_R} \times (Ats(\mathcal{O}) \cap \mathsf{N_C}) \mid B \sqsubseteq_{\mathcal{O}} \exists r.B'\}$ and $L := \{\varepsilon\}$ if $B \sqsubseteq_{\mathcal{O}} A$, and $L := \emptyset$ otherwise. The set of linear inclusions $\mathfrak{I}_{\mathcal{O}}$ captures subsumptions of the form $B \sqsubseteq_{\mathcal{O}} \exists w.A$ in the following sense.

Lemma 4. *Let $\mathcal{O}$ be a flat, cycle-restricted $\mathcal{ELH}_{\mathcal{R}^+}$-ontology.*

1. *If θ is a solution of $\mathfrak{I}_{\mathcal{O}}$, then $w \in \theta(Z_{B \to A})$ implies $B \sqsubseteq_{\mathcal{O}} \exists w.A$.*
2. *If we define $\theta(Z_{B \to A}) := \{w \in \mathsf{N_R}^* \mid B \sqsubseteq_{\mathcal{O}} \exists w.A\}$, then θ is a finite solution of $\mathfrak{I}_{\mathcal{O}}$.*

Example 5. Consider again the system $\Delta_{\Gamma_2,\tau}$ of Example 3, but replace $B \sqsubseteq^? Y$ with $\exists r.B \sqsubseteq^? Y$. The language inclusions corresponding to this constraint are $Y_A \subseteq \{r\}\emptyset$ and $Y_B \subseteq \{r\}\{\varepsilon\}$. The new system again does not have an admissible solution. However, if we consider the ontology $\mathcal{O} = \{B \sqsubseteq A\}$, then there are solutions θ of $\mathfrak{I}_\mathcal{O}$ that satisfy $\varepsilon \in \theta(Z_{B \rightarrow A})$. Thus, if we extend the inclusion $Y_A \subseteq \{r\}\emptyset$ obtained from $\exists r.B \sqsubseteq^? Y$ to $Y_A \subseteq \{r\}\emptyset \cup \{r\}Z_{B \rightarrow A}$, then the new system has a solution θ such that $r \in \theta(Y_A)$ since the other inclusion for Y_A is $Y_A \subseteq \{s\}X_A \cup \{r\}\{\varepsilon\}$. This implies that there is an admissible solution since there are no language inclusions constraining X_A or X_B.

To deal with transitivity axioms, we introduce additional indeterminates of the form $X_{A,t}$, which are constrained by the following linear language inclusions: $\mathfrak{i}_{A,t}(\mathfrak{s}) = X_{A,t} \subseteq f_{A,t}(C_1) \cup \cdots \cup f_{A,t}(C_n) \cup \mathcal{U}_{A,t}(\mathfrak{s})$ where

$$f_{A,t}(C) := \begin{cases} f_A(C') & \text{if } C = \exists r.C' \wedge r \trianglelefteq_\mathcal{O} t, \\ Y_{A,t} & \text{if } C = Y \in \mathsf{N_V}, \\ \emptyset & \text{otherwise.} \end{cases}$$

Intuitively, the difference between $\mathfrak{i}_A^*(\mathfrak{s})$ and $\mathfrak{i}_{A,t}(\mathfrak{s})$ is that, given a particle $\exists t.\exists w.A$ satisfying $\sigma(C_1) \sqcap \cdots \sqcap \sigma(C_n) \sqsubseteq_\mathcal{O} \exists t.\exists w.A$, the right-hand side of $\mathfrak{i}_{A,t}(\mathfrak{s})$ is designed to recognize w instead of tw.

Example 6. Assume that $\Delta_{\Gamma,\tau} = \{\exists r.B \sqsubseteq^? Y, \exists s.X \sqcap \exists r.A \sqsubseteq^? Y, \exists t.B \sqsubseteq^? X\}$ and consider the ontology $\mathcal{O} = \{s \sqsubseteq t, t \sqsubseteq r\}$. Since $\exists r.B \sqsubseteq^? Y$ yields the language inclusion $Y_A \subseteq \{r\}\emptyset$, any solution θ of $\mathfrak{I}_{\Gamma,\tau}^\mathcal{O}$ must satisfy $\theta(Y_A) = \emptyset$. Hence, if θ is admissible, then $\theta(Y_B) \neq \emptyset$. In the presence of $\mathcal{O}$, the new translation also yields the inclusions: $Y_B \subseteq \{r\}\{\varepsilon\}$, $Y_B \subseteq \{s,t,r\}X_B \cup \{r\}\emptyset$ and $X_B \subseteq \{t,r\}\{\varepsilon\}$. Together with $\theta(Y_B) \neq \emptyset$, the first of these inclusions yields $\theta(Y_B) = \{r\}$. Thus, the second inclusion implies that $\varepsilon \in \theta(X_B)$, and thus θ does not solve the third inclusion. Thus, $\mathfrak{I}_{\Gamma,\tau}^\mathcal{O}$ cannot have an admissible solution, corresponding to the fact that $\Delta_{\Gamma,\tau}$ does not have an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier w.r.t. $\mathcal{O}$.

However, if we add the transitivity axiom $t \circ t \sqsubseteq t$ to $\mathcal{O}$, then $\Delta_{\Gamma,\tau}$ has an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier γ with $\gamma(X) = \exists t.B$ and $\gamma(Y) = \exists r.B$ w.r.t. this ontology. The inclusion $\mathfrak{i}_{B,t}(\mathfrak{s}) = X_{B,t} \subseteq \{\varepsilon\}$, obtained from $\mathfrak{s} = \exists t.B \sqsubseteq^? X$, admits solutions θ with $\theta(X_{B,t}) = \{\varepsilon\}$. Hence, if we extend the language inclusion $Y_B \subseteq \{s,t,r\}X_B \cup \{r\}\emptyset$ to the new one $Y_B \subseteq \{s,t,r\}X_B \cup \{r\}\emptyset \cup \{r\}X_{B,t}$, which takes transitivity of t into account, then the new system of language inclusions has an admissible solution with $\theta(Y_B) = \{r\}$ and $\theta(X_B) = \{t\}$, which corresponds to the unifier γ.

Since the definitions of the terms $\mathcal{U}_A(\mathfrak{s})$ and $\mathcal{U}_{A,t}(\mathfrak{s})$ are quite long and technical, we refer to [9] for exact definitions and detailed explanations motivating them. Let $\mathfrak{I}_{\Gamma,\tau}^\mathcal{O}$ be the system of linear language inclusions consisting of $\mathfrak{I}_\mathcal{O}$ and the inclusions $\mathfrak{i}_A^*(\mathfrak{s})$ and $\mathfrak{i}_{A,t}(\mathfrak{s})$ for every subsumption constraint $\mathfrak{s}$ in $\Delta_{\Gamma,\tau}$. Note that the definition of these language inclusions does not only depend on $\Delta_{\Gamma,\tau}$, but also on τ itself (see Definition 4.17 in [9] for the exact definition).

Proposition 2. *Let τ be a subsumption mapping for Γ w.r.t. $\mathcal{O}$. The unification problem $\Delta_{\Gamma,\tau}$ has an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier γ w.r.t. $\mathcal{O}$ that is compatible with τ iff the system of linear language inclusions $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ has a finite, admissible solution.*

The proof of the only-if direction of this proposition makes use of the fact that we can assume without loss of generality that γ is a simple unifier. In fact, this is already taken into account in the definition of $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ (see [9]).

Definition 5. *The $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier γ of $\Delta_{\Gamma,\tau}$ w.r.t. $\mathcal{O}$ is called* simple *if, for all $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq^? X \in \Delta_{\Gamma,\tau}$ and $\exists w.A \in Part(\gamma(X))$, the following holds:*

1. *there exists $i, 1 \leq i \leq n$ such that*
    (a) *C_i is a ground atom and $C_i \sqsubseteq_{\mathcal{O}}^s \exists w.A$, or*
    (b) *$C_i = Y$ is a variable and $\exists w.A \in Part(\gamma(C_i))$, or*
    (c) *$C_i = \exists r.Y$ for a variable Y, $w = sw'$ for some $s \in \mathsf{N_R}$ and $w' \in \mathsf{N_R}^*$, and*
        - *$\exists w'.A \in Part(\gamma(Y))$ and $r \trianglelefteq_{\mathcal{O}} s$, or*
        - *$\exists t.\exists w'.A \in Part(\gamma(Y))$ for a transitive role t s.t. $r \trianglelefteq_{\mathcal{O}} t \trianglelefteq_{\mathcal{O}} s$; or*
2. *There are atoms $At_1, \ldots, At_k, At'$ of $\mathcal{O}$ $(k \geq 0)$ such that:*
    (a) *$At_1 \sqcap \cdots \sqcap At_k \sqsubseteq_{\mathcal{O}} At'$,*
    (b) *for all $\ell \in \{1, \ldots, k\}$, there exists $i \in \{1, \ldots, n\}$ s.t. $\tau(C_i, At_\ell) = 1$, and*
    (c) *$At' \sqsubseteq_{\mathcal{O}}^s \exists w.A$.*

Lemma 5. *If Γ is an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unification problem that is unifiable w.r.t. $\mathcal{O}$, then there exists a subsumption mapping τ for Γ w.r.t. $\mathcal{O}$ such that $\Delta_{\Gamma,\tau}$ has a simple $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unifier σ w.r.t. $\mathcal{O}$ that is compatible with τ.*

3.3   The PSpace Algorithm

Using the results described in the previous two subsections, we can construct an NPSpace decision procedure for unification in $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$ w.r.t. cycle-restricted $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-ontologies. Due to Savitch's theorem [26], this implies that the problem is also in PSpace.

Given an input consisting of an $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-unification problem and a cycle-restricted $\mathcal{ELH}_{\mathcal{R}^+}^{-\top}$-ontology, the algorithm transforms the ontology and the unification problem into flat ones, which we denote as Γ and $\mathcal{O}$. It then proceeds as follows:

1. It guesses a subsumption mapping τ for Γ w.r.t. $\mathcal{O}$. If no such mapping exists, then it fails.
2. It transforms Γ into $\Delta_{\Gamma,\tau}$, and then translates the latter into the set of linear language inclusions $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$.
3. Finally, the algorithm answers "yes" iff $\mathfrak{I}_{\Gamma,\tau}^{\mathcal{O}}$ has a finite, admissible solution.

Flattening can be done in polynomial time and preserves unifiability [5,18]. A mapping $\tau : Ats_{tr}(\Gamma, \mathcal{O}) \times Ats_{tr}(\Gamma, \mathcal{O}) \to \{0, 1\}$ can be guessed in non-deterministic polynomial time, and checking whether it satisfies the properties

of a subsumption mapping (see Definition 3) can clearly also be realized within polynomial space, as can the translations into $\Delta_{\Gamma,\tau}$ and $\mathfrak{I}^{\mathcal{O}}_{\Gamma,\tau}$. Finally, as shown in [18], testing for the existence of a finite, admissible solution of $\mathfrak{I}^{\mathcal{O}}_{\Gamma,\tau}$ can be reduced in polynomial time to checking emptiness of alternating finite automata with ε-transitions, which is a PSpace-complete problem [23]. This shows that the introduced algorithm really is an NPSpace algorithm. Its correctness is an immediate consequence of Propositions 1 and 2. Since PSpace-hardness already holds for the special case of an empty ontology, we thus have shown the following main result of this paper.

Theorem 1. *Deciding unifiability of $\mathcal{ELH}^{-\top}_{\mathcal{R}+}$-unification problems w.r.t. cycle-restricted $\mathcal{ELH}^{-\top}_{\mathcal{R}+}$-ontologies is PSpace-complete.*

4   Conclusion

We have shown that the approach for obtaining a PSpace decision procedure for $\mathcal{EL}^{-\top}$-unification without a background ontology [18] can be extended to unification w.r.t. a cycle-restricted $\mathcal{ELH}_{\mathcal{R}+}$-ontology, i.e., an ontology that may contain general concept inclusions (GCIs) formulated in $\mathcal{EL}^{-\top}$ as well as role inclusion and transitivity axioms, but does not entail a cyclic subsumption of the form $C \sqsubseteq_{\mathcal{O}} \exists r_1.\exists r_2.\cdots\exists r_n.C$ ($n \geq 1$). As explained in the introduction, both considering concept descriptions not containing the top concept $\top$ and considering GCIs and role axioms is motivated by the expressivity employed in the medical ontology SNOMED CT. Dealing with such a background ontology not only makes the approach more complicated due to the more involved characterization of subsumption (see Lemma 1 and Definition 3, compared to the much simpler versions in [18]). It also requires the development of new notions, such as simple unifiers and the extension of the system of linear language inclusions with new indeterminates and corresponding inclusions.

With SNOMED CT in mind, it would be interesting to see whether results on unification (with or without top) can be further extended to ontologies additionally containing so-called right-identity rules, i.e., role axioms of the form $r \circ s \sqsubseteq r$, since they are also needed to get rid of the SEP-triplet encoding mentioned in the introduction. However, extending the characterization of subsumption to this setting is probably a non-trivial problem. From a theoretical point of view, the big open problem is whether one can dispense with the requirement that the ontology must be cycle-restricted. Even for pure $\mathcal{EL}$, decidability of unification w.r.t. unrestricted ontologies is an open problem.

From a practical point of view, the next step is to develop an algorithm that replaces non-deterministic guessing by a more intelligent search procedure. Since the unification problem is PSpace-complete, a polynomial translation of the whole problem into SAT is not possible (unless NP = PSpace). However, one could try to delegate the search for a subsumption mapping to a SAT solver, which interacts with a solver for the additional condition on such a mapping (existence of a finite, admissible solution of $\mathfrak{I}^{\mathcal{O}}_{\Gamma,\tau}$) in an SMT-like fashion [21].

References

1. Baader, F., Binh, N.T., Borgwardt, S., Morawska, B.: Unification in the description logic $\mathcal{EL}$ without the top concept. In: Proceedings of the 23rd International Conference on Automated Deduction (CADE 2011). Lecture Notes in Computer Science, vol. 6803, pp. 70–84. Springer-Verlag (2011)
2. Baader, F., Borgwardt, S., Morawska, B.: Extending unification in $\mathcal{EL}$ towards general tboxes. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012. AAAI Press (2012)
3. Baader, F., Borgwardt, S., Morawska, B.: A goal-oriented algorithm for unification in $\mathcal{ELH}_{\mathcal{R}+}$ w.r.t. cycle-restricted ontologies. In: AI 2012: Advances in Artificial Intelligence – 25th Australasian Joint Conference, Proceedings. Lecture Notes in Computer Science, vol. 7691, pp. 493–504. Springer (2012)
4. Baader, F., Borgwardt, S., Morawska, B.: A goal-oriented algorithm for unification in $\mathcal{ELH}_{R+}$ w.r.t. cycle-restricted ontologies. LTCS-Report 12-05, Chair for Automata Theory, Institute for Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2012). https://doi.org/10.25368/2022.189
5. Baader, F., Borgwardt, S., Morawska, B.: SAT encoding of unification in $\mathcal{ELH}_{\mathcal{R}+}$ w.r.t. cycle-restricted ontologies. In: Automated Reasoning – 6th International Joint Conference, IJCAR 2012, Proceedings. Lecture Notes in Computer Science, vol. 7364, pp. 30–44. Springer (2012)
6. Baader, F., Borgwardt, S., Morawska, B.: SAT encoding of unification in $\mathcal{ELH}_{R+}$ w.r.t. cycle-restricted ontologies. LTCS-Report 12-02, Chair for Automata Theory, Institute for Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2012). https://doi.org/10.25368/2022.186
7. Baader, F., Borgwardt, S., Morawska, B.: Constructing SNOMED CT concepts via disunification. LTCS-Report 17-07, Chair for Automata Theory, Institute for Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2017). https://doi.org/10.25368/2022.237
8. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Kaelbling, L.P., Saffiotti, A. (eds.) IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005. pp. 364–369. Professional Book Center (2005)
9. Baader, F., Fernández Gil, O.: Unification in the Description Logic $\mathcal{ELH}_{\mathcal{R}+}$ without the Top Concept modulo Cycle-Restricted Ontologies (Extended Version). LTCS-Report 24-01, Chair for Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2024), https://lat.inf.tu-dresden.de/research/reports.html
10. Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017)
11. Baader, F., Kapur, D.: Deciding the word problem for ground identities with commutative and extensional symbols. In: Proc. of the 10th Int. Joint Conf. on Automated Reasoning IJCAR 2020, Part I. Lecture Notes in Computer Science, vol. 12166, pp. 163–180. Springer (2020)
12. Baader, F., Marantidis, P., Mottet, A., Okhotin, A.: Extensions of unification modulo ACUI. Math. Struct. Comput. Sci. **30**(6), 597–626 (2020). https://doi.org/10.1017/S0960129519000185
13. Baader, F., Mendez, J., Morawska, B.: UEL: unification solver for the description logic $\mathcal{EL}$ – system description. In: Automated Reasoning - 6th International

Joint Conference, IJCAR 2012, Proceedings. Lecture Notes in Computer Science, vol. 7364, pp. 45–51. Springer (2012)

14. Baader, F., Morawska, B.: Unification in the description logic $\mathcal{EL}$. In: Rewriting Techniques and Applications, 20th International Conference, RTA 2009, Proceedings. Lecture Notes in Computer Science, vol. 5595, pp. 350–364. Springer (2009)

15. Baader, F., Morawska, B.: SAT encoding of unification in $\mathcal{EL}$. In: Logic for Programming, Artificial Intelligence, and Reasoning - 17th International Conference, LPAR-17, Proceedings. Lecture Notes in Computer Science, vol. 6397, pp. 97–111. Springer (2010)

16. Baader, F., Morawska, B.: Unification in the description logic $\mathcal{EL}$. Log. Methods Comput. Sci. **6**(3) (2010)

17. Baader, F., Narendran, P.: Unification of concept terms in description logics. J. Symb. Comput. **31**(3), 277–305 (2001)

18. Baader, F., Nguyen, T.B., Borgwardt, S., Morawska, B.: Deciding unifiability and computing local unifiers in the description logic $\mathcal{EL}$ without top constructor. Notre Dame J. Formal Log. **57**(4), 443–476 (2016)

19. Baader, F., Snyder, W.: Unification theory. In: Robinson, J.A., Voronkov, A. (eds.) Handbook of Automated Reasoning (in 2 volumes), pp. 445–532. Elsevier and MIT Press (2001)

20. Bachmair, L., Ramakrishnan, I.V., Tiwari, A., Vigneron, L.: Congruence closure modulo associativity and commutativity. In: Proc. of the Third International Workshop on Frontiers of Combining Systems (FroCoS 2000). Lecture Notes in Computer Science, vol. 1794, pp. 245–259. Springer (2000)

21. Barrett, C.W., Sebastiani, R., Seshia, S.A., Tinelli, C.: Satisfiability modulo theories. In: Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.) Handbook of Satisfiability - Second Edition, Frontiers in Artificial Intelligence and Applications, vol. 336, pp. 1267–1329. IOS Press (2021). https://doi.org/10.3233/FAIA201017

22. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In: Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004. pp. 298–302. IOS Press (2004)

23. Jiang, T., Ravikumar, B.: A note on the space complexity of some decision problems for finite automata. Inf. Process. Lett. **40**(1), 25–31 (1991)

24. Kapur, D.: Modularity and combination of associative commutative congruence closure algorithms enriched with semantic properties. Log. Methods Comput. Sci. **19**(1) (2023)

25. Narendran, P., Rusinowitch, M.: Any ground associative-commutative theory has a finite canonical system. J. Autom. Reasoning **17**(1), 131–143 (1996)

26. Savitch, W.J.: Relationships between nondeterministic and deterministic tape complexities. J. Comput. Syst. Sci. **4**(2), 177–192 (1970). https://doi.org/10.1016/S0022-0000(70)80006-X

27. Schulz, S., Romacker, M., Hahn, U.: Part-whole reasoning in medical ontologies revisited—introducing SEP triplets into classification-based description logics. In: AMIA 1998, American Medical Informatics Association Annual Symposium. AMIA (1998)

28. Sofronie-Stokkermans, V.: Locality and subsumption testing in $\mathcal{EL}$ and some of its extensions. In: Advances in Modal Logic 7, papers from the Seventh Conference on Advances in Modal Logic. pp. 315–339. College Publications (2008)

29. Suntisrivaraporn, B., Baader, F., Schulz, S., Spackman, K.A.: Replacing SEP-triplets in SNOMED CT using tractable description logic operators. In: 11th Conference on Artificial Intelligence in Medicine, AIME 2007, Proceedings. Lecture Notes in Computer Science, vol. 4594, pp. 287–291. Springer (2007)

Appendix

As mentioned in the main text of the submission, detailed proofs of our results, as well as the exact definition and detailed explanations motivating the construction of the system of linear language inclusions $\mathfrak{I}^{\mathcal{O}}_{\Gamma,\tau}$, are deferred to the accompanying technical report [9]. This report also contains a more detailed description of the previous approaches on which our new approach is based.

   Since the technical report contains a considerably larger number of definitions and auxiliary results, the numbering assigned to definitions, lemmas, etc., in this submission differs from the one given in the technical report. The purpose of this Appendix is to provide a mapping that supports the reader in finding in [9] the missing proofs and other details that had to be left out from the submission due to the space constraints. The following list contains this information.

- Lemma 2 corresponds to Lemma 1.4 in [9].
- The introduction of Section 3 briefly sketches the approaches developed in [16] and [18] for unification in $\mathcal{EL}$ and in $\mathcal{EL}^{-\top}$ w.r.t. the empty ontology, respectively. In the technical report, Section 3 is devoted to provide a more comprehensive summary of these approaches.
- Proposition 1 corresponds to Proposition 4.3 in [9].
- The notion of a simple unifier defined in Definition 5 is introduced in Subsection 4.2.1 of the technical report. Lemma 5 corresponds to Lemma 4.10 in [9].
- The exact definition (Definition 4.17) and detailed explanations motivating the construction of the system $\mathfrak{I}^{\mathcal{O}}_{\Gamma,\tau}$ are presented in Subsection 4.2.2 of the technical report. There, we explain how simple unifiers are taken into account to define the language inclusions in $\mathfrak{I}^{\mathcal{O}}_{\Gamma,\tau}$. In addition, we provide the exact definitions for the terms $\mathcal{U}_A(\mathfrak{s})$ and $\mathcal{U}_{A,t}(\mathfrak{s})$ (immediately above Definition 4.17).
- Lemma 4 corresponds to Lemma 4.13 in [9].
- Proposition 2 corresponds to Proposition 4.18 in [9].
- A more precise analysis of the computational complexity of our algorithm can be found in Subsection 4.3 of the technical report.