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

$$\mathit{ViralInfection} \sqcap \exists \mathit{findingSite}.\mathit{LungStructure},$$

whereas another one models it as

$$\mathit{LungInfection} \sqcap \exists \mathit{causativeAgent}.\mathit{Virus}.$$

Here $\mathit{ViralInfection}$ and $\mathit{LungInfection}$ 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 $\mathit{ViralInfection}$ and $\mathit{LungInfection}$ as variables, and then substituting the first one by

$$\mathit{Infection} \sqcap \exists \mathit{causativeAgent}.\mathit{Virus}$$

and the second one by

$$\mathit{Infection} \sqcap \exists \mathit{findingSite}.\mathit{LungStructure}.$$

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.

DL 2024: 37th International Workshop on Description Logics, June 18–21, 2024, Bergen, Norway
franz.baader@tu-dresden.de (F. Baader); oliver.fernandez@tu-dresden.de (O. Fernández Gil)
0000-0002-4049-221X (F. Baader); 0000-0002-9458-1701 (O. Fernández Gil)

[1]https://www.ihtsdo.org/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]

$$\exists \mathit{findingSite}.\mathit{LungStructure} \sqsubseteq^? \exists \mathit{findingSite}.X,$$

$$\exists \mathit{findingSite}.\mathit{HeartStructure} \sqsubseteq^? \exists \mathit{findingSite}.X.$$

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 $\mathcal{EL}^{-\top}$ 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. An ontology is called cycle-restricted if it does not entail cyclic subsumptions of the form $C \sqsubseteq \exists r_1.\exists r_2. \cdots \exists r_n.C$, where $n \geq 1$ and C is an $\mathcal{EL}$ concept description. Extensions of the result shown in [7] 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

[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].

finding site, however, the presence of the GCIs

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

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

Our new contribution

The purpose of the work we report on here is to investigate the unification problem in the DL $\mathcal{ELH}^{-\top}_{\mathcal{R}^+}$ 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 $\mathcal{ELH}^{-\top}_{\mathcal{R}^+}$-unification problems w.r.t. cycle-restricted $\mathcal{ELH}^{-\top}_{\mathcal{R}^+}$-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 $\mathcal{ELH}^{-\top}_{\mathcal{R}^+}$ 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, $\mathit{UpperBodyStructure}$ would be such a particle. To repair this problem, we first need to show that, in $\mathcal{ELH}^{-\top}_{\mathcal{R}^+}$, 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 NP = 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.

Acknowledgments

This work was partially supported by the German Federal Ministry of Education and Research (BMBF, SCADS22B) and the Saxon State Ministry for Science, Culture and Tourism (SMWK) by funding the competence center for Big Data and AI "ScaDS.AI Dresden/Leipzig". The authors would like to thank Stefan Borgwardt and Francesco Kriegel for helpful discussions on the form of the definitions and axioms used in the current version of SNOMED CT.

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.