Optimal Fixed-Premise Repairs of $\mathcal{EL}$ TBoxes (Extended Abstract)

Francesco Kriegel
Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany
francesco.kriegel@tu-dresden.de

Introduction. Nowadays, data is an important asset — some call it "the oil of the 21st century." Electronic devices have become very popular and with them different forms of data can be created, e.g., text documents, audio files, video files, tables, databases, etc. The knowledge in such data files cannot always be immediately processed by a computing device: image recognition, speech recognition, and natural language processing need to be employed to transform them into a structured form, but also the beliefs and untruths need to be filtered out. Afterwards, the door is wide open to access the knowledge and even reason with it automatically. This has led to the development of formal languages for knowledge representation and reasoning.

One popular family of such languages are Description Logics (DLs) (Baader, Horrocks, Lutz, Sattler, 2017). DL ontologies represent assertional knowledge in the ABox, where objects are assigned to classes or interrelated with each other, and terminological knowledge in the TBox, which describes the class hierarchy as well as implicative rules and integrity constraints. DLs differ in their expressivity and there is always a trade-off to complexity of reasoning. For their well-defined semantics, reasoning produces explainable and deterministic results, e.g., when it comes to deciding whether a statement follows from an ontology. In applications where latency is an important measure, a DL from the $\mathcal{EL}$ family (Baader, Brandt, Lutz, 2005) could be used since common reasoning tasks are decidable in polynomial time. The profile OWL 2 EL (Motik, Cuenca Grau, Horrocks, Wu, Fokoue, Lutz, 2012) of the Web Ontology Language is based on it. The fastest reasoner for the $\mathcal{EL}$ family is ELK (Kazakov, Krötzsch, Simančik, 2014).

When an ontology is used in an application where consequences are drawn from it, the users sometimes find unwanted consequences. These either indicate errors or are privacy-sensitive information. In order to remove such consequences, the ontology needs to be repaired. In the first place, there might be diverse reasons why an ontology is faulty. For instance, at the time of formulating the ontology the knowledge engineers could have had incomplete knowledge, which means that there are counterexamples against consequences that they were not aware of. The ontology could also have been produced by an unsupervised approach from structured data, where unknown counterexamples lead to faulty consequences as well. Furthermore, if the approach is based on machine learning or other imprecise, non-logic-based methods, then the resulting ontology will probably have a large number of faulty consequences.

Related Work. A classical repair is obtained by deleting statements from the ontology (Greiner, Smith, Wilkerson, 1989; Reiter, 1987). In DL ontologies, axiom pinpointing identifies candidates for removal (Baader, Peñaloza, 2010; Baader, Peñaloza, Suntisrivaraporn, 2007; Schlobach, 2005; Schlobach, Cornet, 2003). However, this approach is often too rough as it also erases too many other consequences that might actually be desired. Thus, instead of removing a minimal number of statements, one should rather modify the ontology such that as few consequences as possible are lost, including the unwanted ones. Alternative, less syntax-dependent repair techniques should therefore be developed.

Gentle repairs are obtained by replacing statements with logically weaker ones instead of removing them completely (Baader, Kriegel, Nuradiansyah, Peñaloza, 2018). The framework can be applied to every monotonic logic, and one only needs to specify how statements can be weakened. Although gentle repairs are a big improvement over classical repairs, it is unclear whether they can be optimal in the sense that only a least amount of consequences is lost.

In this regard, we have identified cases where ABoxes can be optimally repaired w.r.t. static TBoxes (Baader, Koopmann, Kriegel, Nuradiansyah, 2021; Baader, Kriegel, 2022). The approach also weakens the statements about the objects in the ABox, but rather compiles the repaired knowledge in one go. In general, there is no single best repair so that initially a (polynomial) selection needs to be made.

For TBoxes, however, no approach that can produce repairs with an optimality guarantee had been available. The reason might be the more complex nature of TBoxes, compared to ABoxes. While assertional knowledge consists of statements about particular objects only, terminological knowledge is expressed as concept inclusions (CIs) $C \sqsubseteq D$ which hold for all objects (if an object satisfies the premise C, then it must also satisfy the conclusion D).

As a novelty, repairs of $\mathcal{EL}$ TBoxes can be obtained by axiomatizing the logical intersection of the input TBox and the theory of a countermodel to the unwanted consequences (Kriegel, 2019). Such a model containing a counterexample can either be manually specified by the knowledge engineer

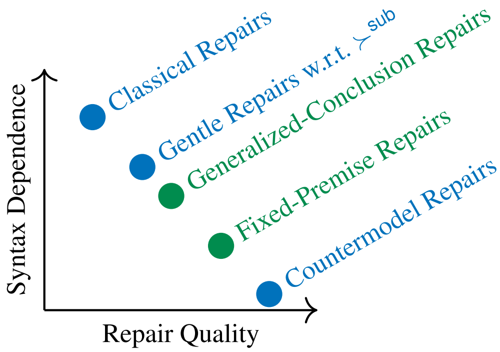

Figure 1: Comparison of $\mathcal{EL}$ TBox Repair Approaches

or be automatically obtained by transforming a canonical model of the TBox (Hieke, Kriegel, Nuradiansyah, 2021). Of course, the quality of the repair depends on the provided countermodel. The axiomatization method is very precise and introduces new premises in the resulting repair if necessary. From a theoretical perspective, this is a clear advantage because a large amount of knowledge can be retained. From a practical perspective, however, this is a disadvantage as the repairs might get considerably larger than the input TBox.

New Contributions. Inspired by the countermodel repairs and improving on the gentle repairs w.r.t. the weakening relation $\succ^{\mathsf{sub}}$, we introduce a framework for computing generalized-conclusion repairs (GC-repairs) of $\mathcal{EL}$ TBoxes, where the premises must not be changed and the conclusions can be generalized. We first devise a canonical construction of such repairs from polynomial-size seeds, and then show that each GC-repair is entailed by an optimal one and that, up to equivalence, the set of all optimal GC-repairs can be computed in exponential time. This complexity cannot be improved as an $\mathcal{EL}$ TBox can have exponentially many optimal GC-repairs and already one of them can be exponential.

In addition, we expand on these results towards fixed-premise repairs (FP-repairs). Unlike GC-repairs, the conclusions of CIs need not be generalizations anymore; only the premises must remain the same and the input TBox must entail each CI in the repair. Thereby even more consequences can be retained. Employing the same seeds as before, we show that every FP-repair is entailed by an optimal one and that the set of all optimal FP-repairs can be computed in exponential time. Also this complexity is not improvable.

An experimental implementation is available, which interacts with the user to construct the seed from which the repair is built. Moreover, we provide new complexity results regarding gentle repairs w.r.t. the weakening relation $\succ^{\mathsf{sub}}$.

Future Prospects. An interesting future task is to combine this approach to optimally repairing TBoxes with the approach to optimally repairing ABoxes. An extension to more expressive DLs is also valuable; ideas from the latest extension of optimal ABox repairs to the DL $\mathcal{ELROI}$ might be helpful. Last, it would be interesting to investigate how the quality of the repairs can be improved if also new premises can be introduced by the repair process.

Acknowledgements. Francesco Kriegel is funded by DFG in Projects 430150274 (ReDLO) and 389792660 (CPEC).

References. This is an abstract of (Kriegel, 2022).

Franz Baader, Sebastian Brandt, Carsten Lutz (2005). Pushing the $\mathcal{EL}$ Envelope. *Proc. of IJCAI*. `http://ijcai.org/Proceedings/05/Papers/0372.pdf`.

Franz Baader, Ian Horrocks, Carsten Lutz, Ulrike Sattler (2017). *An Introduction to Description Logic*. `10.1017/9781139025355`.

Franz Baader, Patrick Koopmann, Francesco Kriegel, Adrian Nuradiansyah (2021). Computing Optimal Repairs of Quantified ABoxes w.r.t. Static $\mathcal{EL}$ TBoxes. *Proc. of CADE*. `10.1007/978-3-030-79876-5_18`.

Franz Baader, Francesco Kriegel (2022). Pushing Optimal ABox Repair from $\mathcal{EL}$ Towards More Expressive Horn-DLs. *Proc. of KR*. `10.24963/kr.2022/3`.

Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, Rafael Peñaloza (2018). Making Repairs in Description Logics More Gentle. *Proc. of KR*. `https://aaai.org/ocs/index.php/KR/KR18/paper/view/18056`.

Franz Baader, Rafael Peñaloza (2010). Axiom Pinpointing in General Tableaux. *J. Log. Comput.* `10.1093/logcom/exn058`.

Franz Baader, Rafael Peñaloza, Boontawee Suntisrivaraporn (2007). Pinpointing in the Description Logic $\mathcal{EL}$. *Proc. of DL*. `http://ceur-ws.org/Vol-250/paper_16.pdf`.

Russell Greiner, Barbara A. Smith, Ralph W. Wilkerson (1989). A Correction to the Algorithm in Reiter's Theory of Diagnosis. *Artif. Intell.* `10.1016/0004-3702(89)90079-9`.

Willi Hieke, Francesco Kriegel, Adrian Nuradiansyah (2021). Repairing $\mathcal{EL}$ TBoxes by Means of Countermodels Obtained by Model Transformation. *Proc. of DL*. `http://ceur-ws.org/Vol-2954/paper-17.pdf`.

Yevgeny Kazakov, Markus Krötzsch, František Simančik (2014). The Incredible ELK – From Polynomial Procedures to Efficient Reasoning with $\mathcal{EL}$ Ontologies. *J. Autom. Reason.* `10.1007/s10817-013-9296-3`.

Francesco Kriegel (2019). *Constructing and Extending Description Logic Ontologies using Methods of Formal Concept Analysis*. Doctoral Thesis. Technische Universität Dresden. `http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa2-360998`.

Francesco Kriegel (2022). Optimal Fixed-Premise Repairs of $\mathcal{EL}$ TBoxes. *Proc. of KI*. `10.1007/978-3-031-15791-2_11`.

Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, Carsten Lutz (2012). *OWL 2 Web Ontology Language Profiles (Second Edition)*. `http://www.w3.org/TR/2012/REC-owl2-profiles-20121211/`.

Raymond Reiter (1987). A Theory of Diagnosis from First Principles. *Artif. Intell.* `10.1016/0004-3702(87)90062-2`.

Stefan Schlobach (2005). Diagnosing Terminologies. *Proc. of AAAI*. `https://www.aaai.org/Papers/AAAI/2005/AAAI05-105.pdf`.

Stefan Schlobach, Ronald Cornet (2003). Non-Standard Reasoning Services for the Debugging of Description Logic Terminologies. *Proc. of IJCAI*. `http://ijcai.org/Proceedings/03/Papers/053.pdf`.