Optimal ABox Repair w.r.t. Static $\mathcal{EL}$ TBoxes: from Quantified ABoxes back to ABoxes

Extended Abstract

Franz Baader, Patrick Koopmann, Francesco Kriegel, Adrian Nuradiansyah

Institute of Theoretical Computer Science, Technische Universität Dresden, 01062 Dresden, Germany

Keywords

Ontology repair, Description logic EL, ABox approximation, Most specific concept, Complexity

Errors in Description Logic (DL) ontologies are often detected when a reasoner computes unwanted consequences. The question is then how to repair the ontology such that the unwanted consequences no longer follow, preferably such that as many of the other consequences as possible are preserved. *ABox repair* deals with the situation where the data (expressed by an ABox) may contain errors, while the schema (expressed by a TBox) is assumed to be correct. For example, the ontology $(\mathcal{T}, \mathcal{A})$, where

$$\mathcal{T} = \{ \ \mathit{Rich} \sqsubseteq \mathit{Famous} \ \}, \qquad \mathcal{A} = \{ \ \mathit{parent}(\mathit{BEN}, \mathit{JERRY}), \ \ \mathit{Rich}(\mathit{JERRY}) \ \},$$

says that rich people are famous and that Ben has a rich parent whose name is Jerry. A *repair request* $\mathcal{R}$ is now a finite set of concept assertions, which each *repair* must not entail. For example, each repair for $\mathcal{R} = \{ \exists \mathit{parent}.(\mathit{Rich} \sqcap \mathit{Famous})(\mathit{BEN}) \}$ must not imply that Ben has a rich and famous parent. Classical approaches for ontology repair are based on removing axioms that are responsible for the undesired entailments [1, 2, 3, 4, 5, 6]. In the present example, we might remove any of the two ABox axioms in order to satisfy the repair request. However, this removes more information than necessary.

Extending on our previous work in [7, 8], our approach in [9] preserves more information than the classical ones since, instead of just removing axioms, we may also introduce anonymous copies of individuals. Specifically, every repair must only logically follow from the input ontology (and need not be a subset as in the classical approaches). The result of the repair process is then a *quantified ABox* (qABox) with existentially quantified variables, which are like anonymous individuals in OWL[1] or nulls in database systems. For the present example, such a repair could be the following qABox:

$$\exists \{y\}. \{ \ \mathit{parent}(\mathit{BEN}, y), \ \ \mathit{Famous}(y), \ \ \mathit{Rich}(\mathit{JERRY}) \ \}.$$

DL 2022: 35th International Workshop on Description Logics, August 7–10, 2022, Haifa, Israel

franz.baader@tu-dresden.de (F. Baader); patrick.koopmann@tu-dresden.de (P. Koopmann); francesco.kriegel@tu-dresden.de (F. Kriegel); adrian.nuradiansyah@tu-dresden.de (A. Nuradiansyah)

0000-0002-4049-221X (F. Baader); 0000-0001-5999-2583 (P. Koopmann); 0000-0003-0219-0330 (F. Kriegel); 0000-0002-9047-7624 (A. Nuradiansyah)

[1] See http://www.w3.org/TR/owl2-direct-semantics/ for the semantics of OWL 2.

In contrast to the classical approaches, it preserves both the information that Ben has a famous parent and that Jerry is rich.

More specifically, in [9] we consider ontologies and repair requests formulated in $\mathcal{EL}$, and develop methods to compute repairs that are *optimal* in the sense that they preserve a maximal subset of some family of Boolean queries. Specifically, we there consider Boolean instance queries (IQ), which are $\mathcal{EL}$ concept assertions, and Boolean conjunctive queries (CQ). IQ-repairs are characterized based on *IQ-entailment*: a qABox $\exists X_1.\mathcal{A}_1$ IQ-entails another qABox $\exists X_2.\mathcal{A}_2$ w.r.t. $\mathcal{T}$ if every Boolean IQ entailed by $\exists X_2.\mathcal{A}_2$ w.r.t. $\mathcal{T}$ is also entailed by $\exists X_1.\mathcal{A}_1$ w.r.t. $\mathcal{T}$. An IQ-repair of a qABox $\exists X.\mathcal{A}$ is a qABox that is IQ-entailed w.r.t. $\mathcal{T}$ by $\exists X.\mathcal{A}$ and does not entail w.r.t. $\mathcal{T}$ any of the the axioms in the repair request. Such a repair is *optimal* if no other repair IQ-entails it without itself being IQ-entailed by it. CQ-repairs can be similarly characterized based on entailed Boolean CQs, but note that CQ-entailment coincides with classical entailment (based on models) because the TBox is fixed. Since each Boolean IQ is also a Boolean CQ but not vice versa, CQ-entailment is stronger than IQ-entailment, and so each CQ-repair is an IQ-repair but the converse does not hold. The repair in our example is both an optimal IQ-repair and an optimal CQ-repair. In general, while optimal IQ-repairs always exist, optimal CQ-repairs are only guaranteed to exist w.r.t. cycle-restricted TBoxes [10].

A downside of the approach in [9] is that, despite being part of the OWL standard, DL systems offer no or only limited support for anonymous individuals. It is thus desirable to find repairs in form of classical ABoxes, which constitutes the research question in our current ESWC 2022 paper [11]. Such (optimal) ABox repairs are defined like (optimal) IQ-repairs, but using classical entailment instead of IQ-entailment and are restricted to being classical $\mathcal{EL}$ ABoxes.

In the above example, we can avoid the quantified variable y if we use the ABox axiom $\exists \mathit{parent}.\mathit{Famous}(\mathit{BEN})$. But this is not always possible. Assume that instead of $\mathcal{T}$, we have the TBox

$$\mathcal{T}' = \{\ \exists \mathit{parent}.\mathit{Rich} \sqsubseteq \mathit{Famous},\quad \mathit{Famous} \equiv \exists \mathit{friend}.\mathit{Famous}\ \}$$

and our repair request is $\mathcal{R}' = \{\ \mathit{Famous}(\mathit{BEN})\ \}$. An optimal IQ-repair for this repair request is the qABox

$$\exists X.\mathcal{A}_1 = \exists\{x, y\}.\{\ \mathit{parent}(\mathit{BEN}, x),\ \ \mathit{Rich}(\mathit{JERRY}),\ \ \mathit{friend}(\mathit{BEN}, y),\ \ \mathit{friend}(y, y)\ \},$$

but no classical ABox is IQ-equivalent to it w.r.t. $\mathcal{T}'$.

The reason is that, due to the cyclic role assertion $\mathit{friend}(y, y)$, the qABox $\exists X.\mathcal{A}_1$ entails $(\exists \mathit{friend}.)^n \top(\mathit{BEN})$ for every $n \geq 0$, which cannot be captured by a classical ABox, and this cycle is also not covered by the TBox. If we replace the last axiom in $\mathcal{T}'$ by $\mathit{Famous} \sqsubseteq \exists \mathit{friend}.\mathit{Famous}$, we obtain $\exists X.\mathcal{A}_2 = \exists X.(\mathcal{A}_1 \cup \{\mathit{Famous}(y)\})$ as optimal IQ-repair. Even though $\mathcal{A}_2$ still contains $\mathit{friend}(y, y)$, it is IQ-equivalent to the classical ABox $\{\ (\exists \mathit{parent}.\top)(\mathit{BEN}),\ \mathit{Rich}(\mathit{JERRY}),\ (\exists \mathit{friend}.\mathit{Famous})(\mathit{BEN})\ \}$ w.r.t. the modified TBox $\mathcal{T}'$. The entailments $(\exists \mathit{friend}.)^n \mathit{Famous}(\mathit{BEN})$ for all $n \geq 0$ are now produced by the the assertion $(\exists \mathit{friend}.\mathit{Famous})(\mathit{BEN})$ together with the last TBox axiom.

These examples demonstrate that optimal repairs in form of classical ABoxes may not always exist, and that it is not obvious to see when they do. As main contribution of our ESWC 2022 paper [11], we show how to decide the existence of such optimal ABox repairs in exponential time, and how to compute in double-exponential time all such repairs in case they exist. There may exist up to exponentially many such repairs, and each repair can be up to double-exponential in size, measured w.r.t. the original (quantified) ABox, the TBox, and the repair request. For data complexity, however, each repair is of at most single-exponential size.

Our approach for showing these results roughly proceeds as follows. First of all, we observe that classical entailment between a qABox and an ABox coincides with so-called IRQ-entailment, which is slightly stronger than IQ-entailment by additionally taking role assertions between named individuals into account. Thus, if we want to characterize the optimal ABox repairs of a (quantified) ABox w.r.t. classical entailment, we can investigate IRQ-repairs instead. In [9] we present two methods for computing optimal IQ-repairs: *canonical IQ-repairs*, which are easier to handle in formal proofs, and *optimized IQ-repairs*, which are IQ-equivalent sub-qABoxes of the canonical repairs that are of exponential size only in the worst case and can thus often be computed within reasonable time bounds. In the present paper, we show that both methods can be used without adaptations to compute all optimal IRQ-repairs. This is by chance since our definition of canonical IQ-repairs does not generate new role assertions between individuals and preserves as many of them as possible, although this is not necessary for IQ-entailment. (Such a result might not hold for other sets of IQ-repairs.) Note that, due to the different entailment relations, not every canonical IQ-repair that is IRQ-optimal is also IQ-optimal.

Each optimal IRQ-repair is a representation of an optimal ABox repair, but we still need to transform it. For this purpose, we introduce the notion of an *optimal ABox approximation* of a given qABox, which is an ABox that entails the same concept assertions and role assertions, i.e., is IRQ-equivalent to the qABox. A given qABox may not have an optimal ABox approximation, but if it does, then this approximation is unique up to equivalence. In particular, if an optimal IRQ-repair has an optimal ABox approximation, then it is an optimal ABox repair. Conversely, every optimal ABox repair can be obtained in this way. Thus, the set of optimal ABox approximations of all optimal IRQ-repairs is the set of all optimal ABox repairs. Contrary to the IQ- and IRQ-repairs, not every ABox repair is entailed by an optimal one. This corresponds to the optimal IRQ-repairs that do not have an optimal ABox approximation. These could still be transformed into an ABox by unfolding up to a fixed role depth. Increasing the role depth then always leads to better ABox repairs.

Furthermore, we investigate the problems of deciding the existence of and of computing optimal ABox approximations. The first step is to transfer the qABox into a specific form, called pre-approximation, which is saturated w.r.t. the TBox and consists of the original role assertions between named individuals and, for each named individual a, a sub-qABox $\mathcal{B}_a$. We prove that the original qABox has an optimal ABox approximation iff all the named individuals a have a *most specific concept* C_a in $\mathcal{B}_a$ w.r.t. the TBox. The optimal ABox approximation is then obtained by replacing each $\mathcal{B}_a$ with $C_a(a)$ in the pre-approximation. We use the results from [12] to test the existence of the most specific concept (msc) in polynomial time and to generate the at most exponentially large msc. The latter implies that an optimal ABox approximation is, up to equivalence, of at most exponential size w.r.t. the size of the qABox. Given that the optimal IRQ-repairs itself are of exponential size in the worst case, this yields the complexity upper bounds for testing the existence and computing optimal ABox repairs mentioned above.

The full paper is published in the proceedings of the 19th Extended Semantic Web Conference (ESWC 2022) [11]. The accompanying technical report with detailed proofs is available in [13].

Acknowledgments

This work was partially supported by the AI competence center ScaDS.AI Dresden/Leipzig and the Deutsche Forschungsgemeinschaft (DFG), Grant 430150274, and Grant 389792660 within TRR 248.

References

[1] S. Schlobach, R. Cornet, Non-standard reasoning services for the debugging of description logic terminologies, in: IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003, Morgan Kaufmann, 2003, pp. 355–362. URL: http://ijcai.org/Proceedings/03/Papers/053.pdf.

[2] B. Parsia, E. Sirin, A. Kalyanpur, Debugging OWL ontologies, in: Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005, ACM, 2005, pp. 633–640. URL: https://doi.org/10.1145/1060745.1060837.

[3] T. A. Meyer, K. Lee, R. Booth, J. Z. Pan, Finding maximally satisfiable terminologies for the description logic $\mathcal{ALC}$, in: Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA, AAAI Press, 2006, pp. 269–274. URL: http://www.aaai.org/Library/AAAI/2006/aaai06-043.php.

[4] S. Schlobach, Z. Huang, R. Cornet, F. van Harmelen, Debugging incoherent terminologies, J. Autom. Reason. 39 (2007) 317–349. URL: https://doi.org/10.1007/s10817-007-9076-z.

[5] A. Kalyanpur, B. Parsia, M. Horridge, E. Sirin, Finding all justifications of OWL DL entailments, in: The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007, volume 4825 of *Lecture Notes in Computer Science*, Springer, 2007, pp. 267–280. URL: https://doi.org/10.1007/978-3-540-76298-0_20.

[6] F. Baader, B. Suntisrivaraporn, Debugging SNOMED CT using axiom pinpointing in the description logic $\mathcal{EL}^{+}$, in: Proceedings of the Third International Conference on Knowledge Representation in Medicine, Phoenix, Arizona, USA, May 31st - June 2nd, 2008, volume 410 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2008. URL: http://ceur-ws.org/Vol-410/Paper01.pdf.

[7] F. Baader, F. Kriegel, A. Nuradiansyah, Privacy-preserving ontology publishing for $\mathcal{EL}$ instance stores, in: Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019, Proceedings, volume 11468 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 323–338. URL: https://doi.org/10.1007/978-3-030-19570-0_21.

[8] F. Baader, F. Kriegel, A. Nuradiansyah, R. Peñaloza, Computing compliant anonymisations of quantified ABoxes w.r.t. $\mathcal{EL}$ policies, in: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part I, volume 12506 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 3–20. URL: https://doi.org/10.1007/978-3-030-62419-4_1.

[9] F. Baader, P. Koopmann, F. Kriegel, A. Nuradiansyah, Computing optimal repairs of quantified ABoxes w.r.t. static $\mathcal{EL}$ TBoxes, in: Automated Deduction - CADE 28 - 28th

International Conference on Automated Deduction, Virtual Event, July 12-15, 2021, Proceedings, volume 12699 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 309–326. URL: https://doi.org/10.1007/978-3-030-79876-5_18.

[10] F. Baader, S. Borgwardt, B. Morawska, SAT Encoding of Unification in $\mathcal{ELH}_{R^+}$ w.r.t. Cycle-Restricted Ontologies, in: Automated Reasoning - 6th International Joint Conference, IJCAR 2012, Manchester, UK, June 26-29, 2012. Proceedings, volume 7364 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 30–44. URL: https://doi.org/10.1007/978-3-642-31365-3_5.

[11] F. Baader, P. Koopmann, F. Kriegel, A. Nuradiansyah, Optimal ABox repair w.r.t. static $\mathcal{EL}$ TBoxes: from quantified ABoxes back to ABoxes, in: 19th Extended Semantic Web Conference, ESWC 2022, Hersonissos, Greece, May 29 – June 2, 2022, Proceedings, volume 13261 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 130–146. URL: https://doi.org/10.1007/978-3-031-06981-9_8.

[12] B. Zarrieß, A. Turhan, Most specific generalizations w.r.t. general $\mathcal{EL}$-TBoxes, in: IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013, IJCAI/AAAI, 2013, pp. 1191–1197. URL: http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6709.

[13] F. Baader, P. Koopmann, F. Kriegel, A. Nuradiansyah, Optimal ABox Repair w.r.t. Static $\mathcal{EL}$ TBoxes: from Quantified ABoxes back to ABoxes (Extended Version), LTCS-Report 22-01, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany, 2022. URL: https://doi.org/10.25368/2022.65.