Evolution of OWL 2 QL Knowledge Bases: From Inexpressibility to Practical Approaches

Evgeny Kharlamov\textsuperscript{1,2}, Dmitriy Zheleznyakov\textsuperscript{3}, Werner Nutt\textsuperscript{4}, and Diego Calvanese\textsuperscript{4}

\textsuperscript{1} Bosch Centre for Artificial Intelligence, Germany, evgeny.kharlamov@de.bosch.com
\textsuperscript{2} University of Oslo, Norway, evgeny.kharlamov@ifi.uio.no
\textsuperscript{3} Ocado Technology, UK, zheleznyakov.dmitry@gmail.com
\textsuperscript{4} Free University of Bozen-Bolzano, Italy, fname.lname@unibz.it

Abstract. Knowledge bases (KBs) are not static entities: new information constantly appears and some of the previous knowledge becomes obsolete. In order to reflect this evolution of knowledge, KBs should be expanded with the new knowledge and contracted from the obsolete one. This problem is well-studied for propositional but much less for first-order KBs. In this work we investigate knowledge expansion and contraction for KBs expressed in OWL 2 QL, a tractable fragment of the Web Ontology Language OWL 2. We start with a novel knowledge evolution framework and natural postulates that evolution should respect, and compare our postulates to the well-established AGM postulates. We then review well-known model and formula-based approaches for expansion and contraction for propositional theories and show how they can be adapted to the case of OWL 2 QL. In particular we show inexpressibility challenges for the former and practical algorithms for the latter approaches.

1 Motivation

Ontology Web Language (OWL) provides excellent mechanisms for representing structured knowledge as knowledge bases (KBs). OWL is the standard ontology language of the Semantic Web. KBs have been successfully used in various applications including Web search \cite{19,12}, and search over KGs \cite{5,39,38}. Medicine \cite{3}, Media \cite{31}, E-commerce \cite{8}, data integration \cite{23,24,20,35} and industrial modelling \cite{22} and analytics \cite{26,24,34,33}. In these and other applications KBs naturally change over time and thus KB management systems should be equipped with services to support KB evolution \cite{13}.

In KB evolution the task is to incorporate new knowledge $\mathcal{N}$ into an existing KB $\mathcal{K}$, or to delete some obsolete knowledge $\mathcal{N}$ from $\mathcal{K}$, in order to take into account changes that occur in the underlying domain of interest \cite{21}. The former evolution task is typically referred to as knowledge expansion and the latter as contraction. In general, the new (resp., obsolete) knowledge is represented by a set of formulas denoting those properties that should be true (resp., false) after the ontology has evolved. In the case where the new knowledge interacts in an undesirable way with the knowledge in the ontology, e.g., by causing the ontology or relevant parts of it to become unsatisfiable, the new knowledge cannot simply be added to the ontology. Instead, suitable changes need to be made in the ontology so as to avoid the undesirable interaction, e.g., by deleting...
parts of the ontology that conflict with the new knowledge. Different choices are possible, corresponding to different semantics for knowledge evolution \[44,41,21,12,36\].

The main two types of semantics that were proposed for the case of propositional knowledge are model-based [41] and formula-based [11]. In model-based semantics the idea is to resolve the undesirable interaction at the level of models of $K$ and $N$. For example, in model-based expansion the result of evolution are those models of $N$ that are minimally distant from the ones of $K$, where a suitable notion of distance needs to be chosen, possibly depending on the application. In formula-based semantics the idea is to do evolution at the level of the deductive closure of the formulae from $K$ and $N$. Since many (possibly counter-intuitive) semantics can be defined within the model or formula-based paradigm, a number of evolution postulates [21,11] have been proposed and they define natural properties a semantics should respect. It is thus common to verify for each evolution semantics whether it satisfies the postulates.

For the case of propositional knowledge, there is a thorough understanding of semantics as well as of computational properties of both expansion and contraction. The situation is however much less clear when it comes to DL KBs, which are decidable first-order logic theories. Differently from the propositional case, in general they admit infinite sets of models and infinite deductive closures. Moreover, going from propositional letters to first-order predicates and interpretations, on the one hand calls for novel postulates underlying the semantics of evolution, and on the other hand broadens the spectrum of possibilities for defining such semantics. A number of attempts have been made to adapt approaches for the evolution of propositional knowledge to the case of DLs, cf. [12,10,36,32]. However, there is no thorough understanding of evolution from the foundational point of view even for DLs with the most favorable computational properties, such as the logics of the OWL 2 QL [7] and EL [6] families, which are at the basis of two tractable fragments of OWL 2.

## 2 Contributions

In this work we address this problem and propose an exhaustive study of evolution for OWL 2 QL. In particular, we address the problem considering three dimensions:

1. knowledge evolution tasks: we study how knowledge can be expanded or contracted;
2. type of evolution semantics: we study model-based and formula-based semantics;
3. evolution granularity: we study when evolution affects the TBox (for terminological knowledge), or the ABox (for assertional knowledge), or both of them.

We provide the following contributions [43]:

- We propose a knowledge expansion and contraction framework that accounts for TBox, ABox, and general KB evolution.
- We propose natural evolution postulates and show how they are related to the well-known AGM postulates [21].
- We show how one can rigorously extend propositional model-based evolution semantics to the first-order case, defining a 5-dimensional space of possible options,
comprising $3 \cdot 2^4$ model-based evolution semantics for DLs that essentially include all previously proposed model-based approaches for DLs. These dimensions are: (1) ABox vs. TBox vs. general evolution; (2) expansion vs. contraction; (3) global vs. local; (4) symbol vs. atom; (5) set inclusion vs. cardinality. For most of these semantics and the case of OWL 2 QL KBs we prove negative expressibility results: in general evolution of OWL 2 QL KBs cannot be expressed as a OWL 2 QL KB.

– We investigate formula-based evolution for OWL 2 QL. In particular, for known formula-based evolution approaches [11] we show intractability of computing evolution results for OWL 2 QL KBs. Moreover, we propose a non-deterministic approach for general KB evolution, which turns out to become deterministic for ABox evolution; for both cases we develop polynomial-time algorithms.

3 Illustration of Contributions

We now exemplify the inexpressibility of evolution results under model-based semantics and then show how the same examples can be solved with formula based semantics.

3.1 Illustration of Inexpressibility

In order to understand why model-based approaches to evolution are problematic for OWL 2 QL recall the following property of the logic. Let $\mathcal{M}$ be a set of interpretations such that there are OWL 2 QL assertions $\varphi, \psi$ such that

- $\mathcal{J} \models \varphi \lor \psi$ for every $\mathcal{J} \in \mathcal{M}$, and
- there are $\mathcal{J}_\varphi, \mathcal{J}_\psi \in \mathcal{M}$ such that $\mathcal{J}_\varphi \not\models \varphi$ and $\mathcal{J}_\psi \not\models \psi$.

Then, there is no OWL 2 QL KB $\mathcal{K}$ such that $\mathcal{M} = \text{Mod}(\mathcal{K})$.

We now illustrate this phenomenon on several evolution scenarios. We will do it on the intuitive level and without referring to concrete model-based semantics. Formal details can be found in [43].

Consider a KB where the structural knowledge is that wives (concept Wife) are exactly those individuals who have husbands (role HasHusband) and that some wives are employed (concept EmpWife). Bachelors (concept Bachelor) cannot be husbands. Priests (concept Priest) are clerics (concept Cleric) and clerics are bachelors. Both clerics and wives are receivers of rent subsidies (concept Renter). We also know that adam and bob are priests, mary is a wife who is employed and her husband is john. Also, carl is a catholic minister (concept Minister).

This knowledge can be expressed in OWL 2 QL by the KB $\mathcal{K}_{ex}$, consisting of the following TBox $\mathcal{T}$ and ABox $\mathcal{A}$:

\[
\mathcal{T} = \{ \text{Wife} \sqsubseteq \exists \text{HasHusband}, \exists \text{HasHusband} \sqsubseteq \text{Wife}, \text{EmpWife} \sqsubseteq \text{Wife}, \text{Bachelor} \sqsubseteq \neg \exists \text{HasHusband}^-, \text{Priest} \sqsubseteq \text{Cleric}, \text{Cleric} \sqsubseteq \text{Bachelor}, \text{Cleric} \sqsubseteq \text{Renter}, \text{Wife} \sqsubseteq \text{Renter} \}
\]

\[
\mathcal{A} = \{ \text{Priest(adam)}, \text{Priest(bob)}, \text{EmpWife(mary)}, \text{HasHusband(mary,john)} \}
\]

\footnote{Note that our proposed model-based semantics can be applied to any description logic.}
In the TBox expansion scenarios the new information $N_T$ can state that wives are not renters anymore or that priests are not renters anymore:

$$N_T = \{\text{Priest} \sqsubseteq \neg \text{Renter}\}, \text{ or } N_T = \{\text{Wife} \sqsubseteq \neg \text{Renter}\}.$$ 

In both cases the inexpressibility phenomenon holds. Indeed, in the first case since priests are not renters anymore, in each model of the evolution result both axioms $\varphi = (\text{Priest} \sqsubseteq \text{Cleric})$, $\psi = (\text{Cleric} \sqsubseteq \text{Renter})$ cannot hold at the same time, but, due to the minimality of change principle, at least one of them should hold. The second case is analogous but with $\varphi = (\text{EmpWife} \sqsubseteq \text{Renter})$ and $\psi = (\text{Wife} \sqsubseteq \text{Renter})$. These two cases hold for different model-based semantics and nicely illustrate that the inexpressibility property affects evolution at the TBox level even when there is a rather simple interaction between atomic concepts such as $A \sqsubseteq B \sqsubseteq C$ or $A \sqsubseteq B$ and $A \sqsubseteq C$. The same effect can be also shown for TBox contraction: instead of adding $\text{Priest} \sqsubseteq \neg \text{Renter}$ to the example KB, one can contract the KB with the axiom $\text{Priest} \sqsubseteq \text{Renter}$ and essentially the same argument for inexpressibility will hold.

In the ABox expansion scenario the new information $N_e$ consider the case when John becomes a priest, that is,

$$N_A = \{\text{Priest}(\text{john})\}.$$ 

The example TBox entails that in this case Mary cannot be the wife of John anymore: John becomes a bachelor who cannot be a husband. In this case one can show that in each model of the evolution result the disjunction of the two ABox axioms $\varphi = \text{Priest}(\text{bob})$ and $\psi = \text{Priest}(\text{adam})$ holds but there are models where one of the two holds but not the other, that is, when either Bob or Adam becomes the new husband of Mary. This again leads to the inexpressibility of the evolution result.

### 3.2 Illustration of Bold Semantics

Given a KB and new knowledge that should be added to the KB, Bold Semantics essentially takes a maximal subset of axioms (entailed) from the KB that together with the new knowledge is satisfiable. There clearly may be more than one such maximal subset and thus Bold Semantics is non-deterministic for OWL 2 QL. On the positive side, computation of evolution results under Bold Semantics can be done in time polynomial in the size of the KB. Moreover, one can show that if the evolution of OWL 2 QL KBs affects only the ABox level, then the evolution result is always unique.

Following our example, evolution with any of the $N_T$ should delete either $\varphi$ or $\psi$ from the original KB, thus there is non-determinism. While, for $N_A$ evolution is unique and it requires to drop from the original KB the fact that Mary is a wife of John but add that she is a wife of someone $\exists \text{HasHusband}(\text{mary})$.

### 3.3 Postulates

Postulates are basic principles that describe the rational behind knowledge evolution. They are typically defined independently from the actual approaches or algorithms to
compute evolution results. Moreover, in order to make sure that a defined approach (algorithm) makes sense one typically verifies whether it confirms some postulates. The classical AGM postulates for knowledge evolution have originally been defined for the case of propositional knowledge. At the same time, the ‘granularity’ of knowledge changes when moving from propositional to Description Logics: the atomic statements of a DL, namely the ABox and TBox axioms, are more complex than the atoms of propositional logic. On the other hand, a set of propositional formulas makes sense, intuitively, if it is satisfiable, while a KB can be satisfiable, but incoherent, that is, one or more concepts are necessarily empty. Therefore, we proposed new postulates for expansion and contraction, to be adopted in the context of evolution on the Semantic Web. Moreover, we showed how our postulates are related to the AGM ones and showed that our evolution operators satisfy the proposed postulates. Due to lack of space we now present only postulates for knowledge expansion.

Let \( K \) be the KB and \( K_e \) the result of its expansion with new knowledge \( N_e \).

**E1:** Expansion should preserve the coherence of the KB, that is, if \( K \) is coherent, then so is \( K_e \).

**E2:** Expansion should entail all new knowledge, that is, \( K_e \models N_e \).

**E3:** Expansion with old information should not affect the KB, that is, if \( K \models N \), then \( K_e \equiv K \).

**E4:** The union of \( N_{1e} \) with the expansion of \( K \) with \( N_{1e} \) implies the expansion of \( K \) with \( N_{1e} \cup N_{2e} \).

**E5:** Expansion should not depend on the syntactical representation of knowledge, that is, if \( K_1 \equiv K_2 \) and \( N_{1e} \equiv N_{2e} \), then \( K_1 \equiv K_2 \).

### 4 Discussion of Contributions

The first important conclusion from our work is that model-based approaches are intrinsically problematic for KB evolution, even in the case of such a lightweight DL as OWL 2 QL. Indeed, recall that OWL 2 QL is not closed under evolution for any of the model-based semantics and thus these semantics are impractical. As a consequence, one has either to search for conceptually different semantics that rely on other principles of ‘composing’ the output set of models constituting the evolution result, or one has to develop natural restrictions on how model-based approaches can ‘compose’ this set. An alternative approach would be to develop approximation techniques that allow one to efficiently capture evolution results.

A second important conclusion is that classical formula-based approaches are too heavyweight from the computational point of view and thus their practicality is questionable. On the other hand, the most conceptually simple model-based semantics such as bold semantics can potentially lead to practical evolution algorithms. However, their practicality requires further empirical evaluation. Finally, we have discussed that the classical evolution postulates that were originally developed for propositional theories are not directly applicable to the case of first-order knowledge since they are blind to some fundamental properties of such knowledge, such as coherency. We have shown how to adapt such postulates to the richer setting considered here, and have analyzed
whether the various model-based and formula-based semantics satisfy the revised postulates.

We believe that our work opens new avenues for research in the area of knowledge evolution, which is an important part of knowledge engineering, since it shows how to lift approaches to knowledge evolution from the propositional to the first-order case. Moreover, we have presented techniques that allow one to prove inexpressibility of model-based evolution, and coNP-hardness of formula-based evolution. We believe that these techniques can be relevant to knowledge management tasks beyond evolution.

5 Future Work

We see several important directions for future work. First, the problem of expressibility in OWL 2 QL is still open for various model-based evolution semantics (see Table 1). These settings are all for ABox expansion and contraction under global model-based semantics. An important research direction is to apply in practice the ideas we developed and, in particular, to implement an ontology evolution system. The system can be based on formula-based approaches and implement Algorithms 1–4 that we proposed. Such system could also be based on approximations of model-based semantics, which are out of the scope of this paper, see, e.g., [30,28,27,29,28,10,17,18]. Then, it would be interesting to conduct an empirical evaluation for various semantics, in order to establish which semantics give more intuitive results from the users’ point of view, and which ABox evolution approaches are more scalable. A further direction to investigate is to identify the minimum extensions of OWL 2 QL that would allow it to capture the results of model-based evolution for OWL 2 QL KBs. For this, one can draw inspiration from the work in [32]. Also, it is still unknown what are minimal DLs that are closed under local model-based evolution, and in general that are well tailored towards model-based approaches. Then, knowledge evolution has important implications to privacy: one should make sure that changes in knowledge do not make violations in access control policies. This is a non-trivial task since, e.g., new knowledge can interact with the old one in such a way that a person without access rights to a particular knowledge can derive such knowledge from this combination [16,14]. Finally, we believe that it is important to develop knowledge evolution techniques where the user has a much better control over the evolution process. For this, one can draw inspiration from previous work, e.g., from [9], where the authors proposed techniques to control what syntactic structures of a given KB cannot be changed by the evolution process, or from [42], where the authors proposed to combine knowledge evolution with models of trust, i.e., the new knowledge in their approach is only partially trusted (note that this scenario inherits the inexpressibility issues of MBAs).

Acknowledgements This work was partially funded by the SIRIUS Centre, Norwegian Research Council project number 237898; by the UNIBZ projects PARCIS and TaDaQua, and by the “European Region Tyrol-South Tyrol-Trentino” (EGTC) under the first call for basic research projects within the Euregio Interregional Project Network IPN12 “Knowledge-Aware Operational Support” (KAOS).
References

2. Google KG. [www.google.co.uk/insidesearch/features/search/knowledge.html](http://www.google.co.uk/insidesearch/features/search/knowledge.html)


