Provenance for the Description Logic $\mathcal{ELH}^r$ (Extended Abstract)

Camille Bourgaux[1], Ana Ozaki[2,4], Rafael Peñaloza[3], and Livia Predoiu[4]

[1]DI ENS, ENS, CNRS, PSL University & Inria, Paris, France
[2]University of Bergen, Norway
[3]University of Milano-Bicocca, Italy
[4]Free University of Bozen-Bolzano, Italy

Abstract. This extended abstract presents our work on provenance for the description logic $\mathcal{ELH}^r$ published at IJCAI 2020 [2].

Important reasoning tasks performed over description logics (DL) ontologies are axiom entailment, i.e. deciding whether a given DL axiom follows from the ontology; and query answering. In many settings it is crucial to know *how* a consequence—e.g. an axiom or a query—has been derived from the ontology. In the database community, provenance has been studied for nearly 30 years [5, 3] and gained traction when the connection to semirings, so called *provenance semirings* [6, 7] was discovered. *Provenance semirings* serve as an abstract algebraic tool to record and track provenance information; that is, to keep track of the specific database tuples used for deriving the query, and of the way they have been processed in the derivation. Besides explaining a query answer, provenance has many applications like: computing the probability or the degree of confidence of an answer, counting the different ways of producing an answer, handling authorship, data clearance, or user preferences. Semiring provenance has drawn interest beyond relational databases, and in particular, it has recently been considered for ontology-mediated query answering [1] and for ontology-based data access [4], a setting where a database is enriched with an ontology and mappings between them. In the latter, the ontology axioms are annotated with *provenance variables*. Queries are then annotated with *provenance polynomials* that express their provenance information.

Example 1. Consider the facts $\mathsf{mayor}(\mathsf{Venice}, \mathsf{Brugnaro})$ and $\mathsf{mayor}(\mathsf{Venice}, \mathsf{Orsoni})$, stating that Venice has mayors Brugnaro and Orsoni, annotated respectively with provenance information v_1 and v_2, and the DL axiom $\mathsf{ran}(\mathsf{mayor}) \sqsubseteq \mathsf{Mayor}$, expressing that the range of the role mayor is the concept Mayor, annotated with v_3. The query $\exists x.\mathsf{Mayor}(x)$ asks if there is someone who is a mayor. The answer is *yes* and it can be derived using $\mathsf{ran}(\mathsf{mayor}) \sqsubseteq \mathsf{Mayor}$ together with any of the two facts, interpreting x by $\mathsf{Brugnaro}$ or Orsoni. This is expressed by the provenance polynomial $v_1 \times v_3 + v_2 \times v_3$. Intuitively, $\times$ expresses the joint use of axioms in a derivation path of the query, and $+$ the alternative derivations.

An important contribution of [4] was a semantic definition of provenance for annotated DL-Lite$_\mathcal{R}$ ontologies—which contains axioms of the form (α, v) where α is a DL axiom and v a provenance variable—based on annotated models. An annotated interpretation $\mathcal{I}$ interprets concepts (resp. roles) by sets of pairs of a domain element and a provenance monomial (resp. triples of two domain elements and a provenance monomial). It satisfies e.g. an annotated concept assertion $(A(a), v)$ if $(a^\mathcal{I}, v^\mathcal{I}) \in A^\mathcal{I}$, and an annotated general concept inclusion (GCI) $(C \sqsubseteq D, v)$ if $(d, m^\mathcal{I}) \in C^\mathcal{I}$ implies that $(d, (v \times m)^\mathcal{I}) \in D^\mathcal{I}$. Annotated models and entailment of DL axioms annotated with provenance monomials are then defined as expected. Finally, entailment of conjunctive queries annotated with sums of monomials is defined by considering the matches of the query (extended with binary and ternary predicates to handle provenance information) in the annotated models of the ontology. However, [4] only considered DL-Lite$_\mathcal{R}$, which has the particularity of not allowing for conjunctions or qualified role restrictions in the TBox axioms. Our first contribution is to adapt the provenance semantics of [4] for the $\mathcal{ELH}^r$ variant of the lightweight DL $\mathcal{EL}$, extending the semantics to those $\mathcal{ELH}^r$ axioms that do not occur in DL-Lite$_\mathcal{R}$. It turns out that handling the conjunction allowed in $\mathcal{ELH}^r$ axioms is not trivial. To obtain models from which we can derive meaningful provenance-annotated consequences, we adopt $\times$-idempotent semirings and a syntactic restriction on $\mathcal{ELH}^r$ that forbids conjunctions and qualified restriction of a role to appear in the right-hand side of GCIs (preserving the expressivity of full $\mathcal{ELH}^r$ when annotations are not considered). We then devise methods to handle annotated ontologies in this context.

1. We present a completion algorithm for annotated $\mathcal{ELH}^r$ ontologies that computes all axioms annotated with provenance monomials that follow from the ontology in exponential time. We show that it allows us to solve annotated axiom entailment and instance queries in $\mathcal{ELH}^r$ in polynomial time in the size of the ontology and polynomial space in the size of the provenance polynomial.
2. We also consider the problem of computing the set of relevant provenance variables for the entailment of an axiom (or instance query) α from an annotated ontology $\mathcal{O}$, i.e., the set of variables v such that $\mathcal{O} \models (\alpha, v \times m)$ for some monomial m. We show that this can be done in polynomial time, using an adaptation of the completion algorithm.
3. Finally, we investigate conjunctive query answering. The query answering methods developed in [4] cannot be extended to $\mathcal{ELH}^r$ since they rely on the FO-rewritability of conjunctive queries in DL-Lite$_\mathcal{R}$, a property that does not hold for $\mathcal{ELH}^r$. Therefore, we adapt the combined approach for query answering in $\mathcal{EL}$ [8] to provenance-annotated $\mathcal{ELH}^r$ ontologies. We define a finite canonical model and a rewriting of the query such that the ontology entails the original annotated query if and only if the canonical model satisfies the rewritten query.

References

1. Camille Bourgaux and Ana Ozaki. Querying attributed DL-Lite ontologies using provenance semirings. In *AAAI*, 2019.
2. Camille Bourgaux, Ana Ozaki, Rafael Peñaloza, and Livia Predoiu. Provenance for the description logic $\mathcal{ELH}^r$. In *IJCAI*, 2020.
3. Peter Buneman. The providence of provenance. In *BNCOD*, 2013.
4. Diego Calvanese, Davide Lanti, Ana Ozaki, Rafael Peñaloza, and Guohui Xiao. Enriching ontology-based data access with provenance. In *IJCAI*, 2019.
5. James Cheney, Laura Chiticariu, and Wang Chiew Tan. Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, 1(4), 2009.
6. Todd J. Green, Gregory Karvounarakis, and Val Tannen. Provenance semirings. In *PODS*, 2007.
7. Todd J. Green and Val Tannen. The semiring framework for database provenance. In *PODS*, 2017.
8. Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *IJCAI*, 2009.