Querying Attributed DL-Lite Ontologies Using Provenance Semirings (Extended Abstract)

Camille Bourgaux and Ana Ozaki

DI ENS, CNRS, ENS, PSL University & Inria, France
KRDB Research Centre, Free University of Bozen-Bolzano, Italy

Knowledge graphs often enrich data with contextual information (such as temporal validity or provenance of an assertion) in the format of annotations, which can be seen as *attribute-value pairs*. Recently, attributed description logics have been proposed to bridge the gap between the data format present in knowledge graphs and classical description logics [4, 3, 5]. The mentioned works analyse the complexity of the satisfiability problem for various attributed description logics. This extended abstract gives an overview of our work [1], published at AAAI 2019, on query answering in attributed description logics using provenance semirings. After investigating satisfiability and query answering in attributed DL-Lite$_\mathcal{R}$, we turn our attention to *provenance information*, one of the most common types of annotation in knowledge graphs. Indeed, since knowledge graphs often integrate data from multiple sources, one may not be only interested in obtaining query results but also in establishing levels of trust, or determining authorship, among others [7]. To deal with provenance information, we propose a new semantics based on provenance semirings, as first introduced in database theory [2] (and also recently studied in an ontology-based data access setting [6]).

Attributed DL-Lite$_\mathcal{R}$ Attributed DLs are defined over the usual DL signature with countable sets of *concept names* $\mathsf{N_C}$, *role names* $\mathsf{N_R}$, and *individual names* $\mathsf{N_I}$. We consider an additional set $\mathsf{N_U}$ of *set variables* and a set $\mathsf{N_V}$ of *object variables*. Annotation sets are defined as finite binary relations, understood as sets of attribute-value pairs. Attributes and values refer to domain elements and are syntactically denoted by individual names. To describe annotation sets, we use a set $\mathbf{S}$ of *specifiers*, which can be either *set variables*; *closed specifiers* $[a_1\!:\!v_1, \ldots, a_n\!:\!v_n]$; or *open specifiers* $\lfloor a_1\!:\!v_1, \ldots, a_n\!:\!v_n \rfloor$, where $a_i \in \mathsf{N_I}$ and v_i is either an individual name in $\mathsf{N_I}$, an object variable in $\mathsf{N_V}$, or an expression of the form $X.a$, with X a set variable in $\mathsf{N_U}$ and a an individual name in $\mathsf{N_I}$. We use $X.a$ to refer to the (finite, possibly empty) set of all values of attribute a in an annotation set X. Intuitively, closed specifiers define specific annotation sets whereas open specifiers merely provide lower bounds [3]. A *ground specifier* is a closed or open specifier that only contains individual names. A DL-Lite$_@^\mathcal{R}$ *role* (resp. *concept*) *assertion* is an expression $R(a, b)@S$ (resp. $A(a)@S$), with $R \in \mathsf{N_R}$ (resp. $A \in \mathsf{N_C}$), $a, b \in \mathsf{N_I}$, and $S \in \mathbf{S}$ a ground closed specifier. DL-Lite$_@^\mathcal{R}$ *role* and *concept inclusions* are of the form $X\!:\!S\ (P \sqsubseteq Q)$ and $X\!:\!S\ (B \sqsubseteq C)$ respectively, where $X \in \mathsf{N_U}$, $S \in \mathbf{S}$ is a closed or open specifier, and P, Q and B, C are respectively role and concept expressions defined by the following syntax, where $A \in \mathsf{N_C}$, $R \in \mathsf{N_R}$ and $S \in \mathbf{S}$:

$$P ::= R@S \mid R^-@S, \quad Q ::= P \mid \neg P, \quad B ::= A@S \mid \exists P, \quad C ::= B \mid \neg B.$$

We require that all variables are *safe*, which intuitively means that if they appear on the right side of an inclusion then they should also be associated to the concept on the left.

Example 1. In DL-Lite$_@^\mathcal{R}$, we can express that those who are married (role spouse) to someone are married (concept Married), and that this fact is associated with the same sources from which the information has been extracted (attribute src): $\exists$spouse@$X \sqsubseteq$ Married@$\lfloor$src : X.src$\rfloor$. The assertion spouse(gabor, ryan)@[src: s_1, src: s_2] states that Gabor is married to Ryan and it is annotated with the sources of this information.

We similarly define attributed conjunctive queries by associating concept and role names with specifiers. The formal definition of the semantics is given in [1].

Example 2. The query $\exists x$ Married(gabor)@$\lfloor$year : $x\rfloor \wedge$ Married(taylor)@$\lfloor$year : $x\rfloor$ holds if Gabor and Taylor were married (to someone) on the same year.

We establish complexity results for the satisfiability and query entailment problems.

Theorem 3. *In DL-Lite$_@^\mathcal{R}$, satisfiability and query entailment are* PSPACE-*complete. When the ontology does not contain variables, query entailment is* NP-*complete.*

We now define a new semantics for dealing with provenance, which is based on semirings.

Provenance semirings-based semantics We study query entailment in attributed DL-Lite using a provenance semiring-based semantics. We represent provenance with the *positive algebra provenance semiring* for N_I, defined as the commutative semiring of polynomials with variables in N_I and coefficients from $\mathbb{N}$, with operations defined as usual: $\mathcal{K} = (\mathbb{N}[N_I], +, \times, 0, 1)$ [2]. We denote by N_P the set of polynomials of $\mathcal{K}$ and by N_S the subset of N_P containing the sums of the commutative monoid $(\mathbb{N}[N_I], +, 0)$. We allow to use *provenance polynomials* from N_P as values in a specifier associated to the *whole query* and *provenance sums* from N_S as values in specifiers occurring *in the ontology* or associated with query atoms. Intuitively, $+$ indicates *alternative* use of the data while $\times$ indicates *join* use of the data. For instance, if the result of a query over an ontology can be obtained from source s_3 together with any of s_1, s_2 then the provenance polynomial is: $(s_1 + s_2) \times s_3$. The main idea is that now we allow the whole query to be annotated with a provenance polynomial and such query is entailed if the polynomial represents the alternative and join uses of the data. More specifically, we introduce *semiring attributed queries*, which are attributed queries where we associate a specifier to the whole query [1].

Example 4. The query $($Married$(a) \wedge$ Married$(b))$@$\lfloor$src: $s_1 \times (s_2 + s_3)$, cls: pub$\times$cnf$\rfloor$ is entailed by the assertions Married(a)@[src: s_1, cls: pub], Married(b)@[src: s_2, cls: cnf], and Married(b)@[src: s_3, cls: cnf]. The fact that a and b are both married is obtained by combining s_1 with s_2 or s_3, and by having access to public and confidential information. Sums may also appear in concept inclusions, e.g., $X :\lfloor$src : $s_1 + s_2\rfloor(\exists$spouse@$X \sqsubseteq$ Married@$X)$, which requires that the fact that someone has a spouse has to be associated both with s_1 and with s_2 to conclude that this person is married.

We establish complexity results for this variant of attributed DL-Lite, which we call DL-Lite$_{@,\mathbb{K}}^\mathcal{R}$. The salient results of our investigation are the EXPTIME-hardness for the satisfiability problem, and the NP-completeness of query entailment for a restricted class of ontologies, called *simple*, that allows for inclusions among atomic concepts or roles of the form E_1@$S \sqsubseteq E_2$@T where S and T are ground specifiers.

Theorem 5. *In DL-Lite$_{@,\mathbb{K}}^\mathcal{R}$, satisfiability is* EXPTIME-*hard and in* 2EXPTIME. *When the ontology is simple, query entailment is* NP-*complete.*

References

1. Camille Bourgaux and Ana Ozaki. Querying attributed DL-Lite ontologies using provenance semirings. In *Proceedings of AAAI*, 2019.
2. Todd J. Green, Gregory Karvounarakis, and Val Tannen. Provenance semirings. In *Proceedings of PODS*, 2007.
3. Markus Krötzsch, Maximilian Marx, Ana Ozaki, and Veronika Thost. Attributed description logics: Ontologies for knowledge graphs. In *Proceedings of ISWC*, 2017.
4. Markus Krötzsch, Maximilian Marx, Ana Ozaki, and Veronika Thost. Attributed description logics: Reasoning on knowledge graphs. In *Proceedings of IJCAI*, 2018.
5. Ana Ozaki, Markus Krötzsch, and Sebastian Rudolph. Happy ever after: Temporally attributed description logics. In *Proceedings of DL*, 2018.
6. Ana Ozaki and Rafael Peñaloza. Provenance in ontology-based data access. In *Proceedings of DL*, 2018.
7. Pierre Senellart. Provenance and probabilities in relational databases. *SIGMOD Record*, 46(4):5–15, 2017.