A Unified Logical Framework for Rules (and Queries) with Ontologies - *position paper* -

Enrico Franconi Sergio Tessaris

Faculty of Computer Science, Free University of Bozen-Bolzano, Italy
lastname@inf.unibz.it

Abstract

In this position paper we briefly present a semantic framework for investigating the problem of combining ontology and rule languages. We also show how rules are strictly related to queries. We claim that any choice of rule language for the semantic web should clearly define its semantics.

1 Introduction

The need for integrating rules within the Semantic Web framework was clear since the early developments. However, up to the last few years, the research community focused its efforts on the design of the so called *Ontology Layer*. Nowadays, this layer is fairly mature in the form of Description Logics based languages such as OWL-Lite and OWL-DL, which are now among W3C recommendations.

One of the key features of SW ontology languages development is the attention to the computational properties of the main reasoning tasks. In particular, decidability is seen as one of the characteristics which should be preserved by these languages. This constraint led to the restriction of the expressivity of ontology language which can be heavy for certain applications (e.g. Web Services, or integration of information systems). The problem increasing the expressivity of SW ontology languages over the established Ontology Layer, together with the need of providing powerful query languages, directed the research towards the investigation of the possibility of combining OWL languages with Rules based languages.

In recent years, more research has been devoted towards the integration of different sorts of rule based languages on top of the ontology layer provided by the OWL languages and in more general terms on top of a generic DL, and this work already produced some proposals for extending OWL languages. However, these proposals comes from different research communities, and often are difficult to compare because of the diverse underlying semantic assumptions.

In our works we have provided an unifying framework in which the existing (and future) proposals about rule extended ontology languages can be compared. Moreover, we present a thorough analysis of the main contributions, with a particular attention to their expressive power and restrictions to guarantee the decidability of key inference problems. By using our framework, we show that – under the appropriate restrictions – there are strong correspondences among the proposals. This enable us to isolate interesting fragments of the proposed languages in which we can compare the reasoning abilities.

We reckon that, since the early 90s, the Description Logics community produced several important results w.r.t. the problem of integrating DL languages and rules. For this reason we do not restrict our analysis to proposals in the context of Semantic Web. On the contrary, we show that a careful analysis of this body of work provides a valuable reference to explore the borders of expressivity and tractability of the combination of the two kinds of language.

In our work we identify three different approaches: the axiom-based approach, the logic programming approach, and the autoepistemic approach. We provide an exact characterisation of the three approaches, together with a correspondence among relevant fragments in the three cases. It is important to note that the three different semantics of rule languages lead to different behaviours of the inference in presence of ontologies. The details and all the references can be found in [Franconi and Tessaris, 2004].

Moreover, we turn our attention at the problem of querying knowledge represented by means of an ontology web language. We show that there is a strong connection between rules and queries, and that our

This work has been partially supported by the EU projects Sewasie, KnowledgeWeb, and Interop.

This work has been partially supported by the EU projects Sewasie, KnowledgeWeb, and Interop.

framework is able to capture this fundamental aspect of reasoning in the Semantic Web. At the end of this position paper, we also recall our work to interoperate description logic based ontology languages with RDF [Franconi *et al.*, 2004].

Our work on a common framework is directed to provide the Semantic Web community a foundational tool which can be the basis for the discussion towards a common rule language for the Semantic Web with an agreed upon semantics.

2 Rule-extended Knowledge Bases

Let us consider a first-order function-free language with signature $\mathcal{A}$, and a description logic (DL) knowledge base Σ with signature subset of $\mathcal{A}$.

In this paper we do not introduce any particular DL formalism. In our context, DL individuals correspond to constant symbols, DL atomic concepts and roles (and features) are unary and binary predicates in the case of a classical DL or a OWL language, and DL atomic n-ary relations correspond to predicates of arity n in the case of a $\mathcal{DLR}$-like DL. Note that description logics with concrete data-types (such as OWL-Lite) are allowed as well.

A *term* is any constant in $\mathcal{A}$ or a variable symbol. If R is a predicate symbol of arity n and $t_1, \ldots, t_n$ are terms, $R(t_1, \ldots, t_n)$ is an *atom*, and an atom $R(t_1, \ldots, t_n)$ or a negated atom $\neg R(t_1, \ldots, t_n)$ are *literals*. A *ground* literal is a literal involving only constant terms. A set of ground literals is *consistent* if it does not contain an atom and its negation. If l is a literal, l or *not* l are *NAF-literals* (negation as failure literals). DL atoms, DL literals, and DL NAF-literals are atoms, literals, and NAF-literals whose predicates belong to the DL signature. A *rule* r may be of the forms:

$$\begin{array}{rcll} h_1 \wedge \ldots \wedge h_\ell & \leftarrow & b_1 \wedge \ldots \wedge b_m & \text{(classical rule)} \\ h_1 & :- & b_1 \wedge \ldots \wedge b_m \wedge & \text{(lp-rule)} \\ & & \mathit{not}\ b_{m+1} \wedge \ldots \wedge \mathit{not}\ b_n & \\ h_1 \wedge \ldots \wedge h_\ell & \Leftarrow & b_1 \wedge \ldots \wedge b_m & \text{(autoepistemic rule)} \end{array}$$

where $h_1, \ldots, h_\ell, b_1, \ldots, b_n$ are literals. Given a rule r, we denote by $H(r)$ the set $\{h_1, \ldots, h_\ell\}$ of *head* literals, by $B(r)$ the set of *body* literals $\{b_1, \ldots, b_n\}$, by $B^+(r)$ the set of *NAF-free* body literals $\{b_1, \ldots, b_m\}$, and by $B^-(r)$ the set of *NAF-negated* body literals $\{b_{m+1}, \ldots, b_n\}$. We denote by $\mathsf{vars}(\{l_1, \ldots, l_n\})$ the set of variables appearing in the literals $\{l_1, \ldots, l_n\}$. The *distinguished variables* of a rule r are the variables that appears both in the head and in the body of the rule, i.e., $D(r) = \mathsf{vars}(H(r)) \cap \mathsf{vars}(B(r))$. A *ground rule* is a rule involving only ground literals. A rule is *safe* if all the variables in the head of the rule are distinguished. A *DL rule* is a rule with only DL literals. A set of literals is *tree-shaped* if its co-reference graph is acyclic; a co-reference graph includes literals and variables as nodes, and labelled edges indicate the positional presence of a variable in a literal. An *atomic* rule is a rule having a single literal in the head. A set of rules is *acyclic* if they are atomic and no head literal transitively depends on itself; a head literal h directly depends on a literal l if there is an atomic rule r with head h and with l part of the body $B(r)$. A set of rules is a *view* set of rules if each rule is atomic and no head literal belongs to the DL signature. A *rule-extended knowledge base* $\langle \Sigma, \mathcal{R} \rangle$ consists of a DL knowledge base Σ and a finite set $\mathcal{R}$ of rules.

3 The axiom-based approach

Let us consider a rule-extended knowledge base $\langle \Sigma, \mathcal{R} \rangle$ restricted to only classical rules.

Let I_Σ be a model of the description logics knowledge base Σ, i.e. $I_\Sigma \models \Sigma$. I is a model of $\langle \Sigma, \mathcal{R} \rangle$, written $I \models \langle \Sigma, \mathcal{R} \rangle$, if and only if I extends I_Σ with the interpretation of the non-DL predicates, and for each rule $r \in \mathcal{R}$ then

$$I \models \forall \mathbf{x}, \mathbf{y}. \exists \mathbf{z}. \left(\bigwedge B(r) \rightarrow \bigwedge H(r) \right)$$

where $\mathbf{x}$ are the distinguished variables of the rule $D(r)$, $\mathbf{y}$ are the non distinguished variables of the body $(\mathsf{vars}(B(r)) \setminus D(r))$, and $\mathbf{z}$ are the non distinguished variables of the head $(\mathsf{vars}(H(r)) \setminus D(r))$.

Let us define now the notion of logical implication of a ground literal l given a rule extended knowledge base: $\langle \Sigma, \mathcal{R} \rangle \models l$ if and only if $I \models l$ whenever $I \models \langle \Sigma, \mathcal{R} \rangle$. Note that the problems of DL concept subsumption and DL instance checking, and the problem of predicate inclusion (also called *query containment*) are all reducible to the problem of logical implication of a ground literal. Logical implication in this framework is undecidable. Logical implication in an axiom-based rule extended knowledge base remains undecidable even in the case of atomic negation-free safe DL rules with a DL having just the universal role constructor $\forall R.\, C$. Note that logical implication in an axiom-based rule extended knowledge base even with an empty TBox in Σ is undecidable.

The SWRL proposal [Horrocks and Patel-Schneider, 2004] can be considered as a special case of the axiom-based approach presented above. SWRL uses OWL-DL or OWL-Lite as the underlying description logics knowledge base language (which admits data types), but it restricts the rule language to safe rules and without negated atomic roles. From the point of view of the syntax, SWRL rules are an extension of the abstract syntax for OWL DL and OWL Lite; SWRL rules are given an XML syntax based on the OWL XML presentation syntax; and a mapping from SWRL rules to RDF graphs is given based on the OWL RDF/XML exchange syntax. Logical implication in SWRL is still undecidable.

In order to recover decidability of the axiom-based approach, we should reduce the expressivity of the

rules or of the description logic language; for the list of all the decidable sub-cases, see [Franconi and Tessaris, 2004].

4 The DL-Log approach

Let us consider a rule-extended knowledge base $\langle \Sigma, \mathcal{R} \rangle$ where $\mathcal{R}$ is restricted to be a view set of lp-rules $\mathcal{P}$ (called *program*).

The *non-DL Herbrand base* of the program $\mathcal{P}$, denoted by $\mathcal{HB}_{\mathcal{P}^-}$, is the set of all ground literals obtained by considering all the non-DL predicates in $\mathcal{P}$ and all the constant symbols from $\mathcal{A}$. An *interpretation* I wrt $\mathcal{P}$ is a consistent subset of $\mathcal{HB}_{\mathcal{P}^-}$. We say I is a *model* of a ground literal l wrt the knowledge base Σ, denoted $I \models_\Sigma l$, if and only if

- $l \in I$, when $l \in \mathcal{HB}_{\mathcal{P}^-}$
- $\Sigma \models l$, when l is a DL literal

We say that I is a model of a ground rule r, written $I \models_\Sigma r$, if and only if $I \models_\Sigma H(r)$ whenever $I \models_\Sigma b$ for all $b \in B^+(r)$, and $I \not\models_\Sigma b$ for all $b \in B^-(r)$. We denote with $\mathsf{ground}(\mathcal{P})$ the set of rules corresponding to the grounding of $\mathcal{P}$ with the constant symbols from $\mathcal{A}$. We say that I is a model of a rule-extended knowledge base $\langle \Sigma, \mathcal{P} \rangle$ if and only if $I \models_\Sigma r$ for all rules $r \in \mathsf{ground}(\mathcal{P})$; this is written as $I \models \langle \Sigma, \mathcal{P} \rangle$.

Let us define now the notion of logical implication of a ground literal l given a rule extended knowledge base: $\langle \Sigma, \mathcal{P} \rangle \models l$ if and only if $I \models_\Sigma l$ whenever $I \models \langle \Sigma, \mathcal{P} \rangle$. In the case of a NAF-free program, as well in the case of a program with stratified NAF negation, it is possible to adapt the standard results of datalog, which say that in these cases the logical implication can be reduced to model checking in the (canonical) minimal model. So, if $I_m^{\mathcal{P}}$ is the minimal model of a NAF-free or stratified program $\mathcal{P}$, then $\langle \Sigma, \mathcal{P} \rangle \models l$ if and only if $I_m^{\mathcal{P}} \models_\Sigma l$.

Reasoning in the DL-Log approach is decidable, and the precise complexity bounds have been devised for OWL-Lite and OWL-DL, see [Franconi and Tessaris, 2004].

In [Grosof *et al.*, 2003] the DLP approach is introduced. In this work it is shown how to *encode* the reasoning problem of a DL into a pure logic programming setting, i.e., into a rule extended knowledge base with a Σ without TBox. In the case of DLP, this is accomplished by encoding a severely restricted DL into a NAF-free negation-free DL program.

5 The autoepistemic approach

Let us consider a rule-extended knowledge base restricted to autoepistemic rules.

Let I_Σ be a model, over the non empty domain Δ, of the description logics knowledge base Σ, i.e. $I_\Sigma \models \Sigma$. Let's define a variable assignment α in the usual way as a function from variable symbols to elements of Δ. A model of $\langle \Sigma, \mathcal{R} \rangle$ is a non empty set M of interpretations I, each one extending a DL model I_Σ with some interpretation of the non-DL predicates, such that for each rule r and for each assignment α for the distinguished variables of r the following holds:

$$(\forall I \in M.\ I, \alpha \models \exists \mathbf{x}.\ \bigwedge B(r)) \rightarrow (\forall I \in M.\ I, \alpha \models \exists \mathbf{y}.\ \bigwedge H(r))$$

where $\mathbf{x}$ are the non distinguished variables of the body ($\mathsf{vars}(B(r)) \setminus D(r)$), and $\mathbf{y}$ are the non distinguished variables of the head ($\mathsf{vars}(H(r)) \setminus D(r)$).

Let us define now the notion of logical implication of a ground literal l given a rule extended knowledge base: $\langle \Sigma, \mathcal{R} \rangle \models l$ if and only if

$$\forall M.\ (M \models \langle \Sigma, \mathcal{R} \rangle) \rightarrow \forall I \in M.\ (I \models l)$$

Logical implication in the autoepistemic approach is decidable in some restricted case; see [Franconi and Tessaris, 2004].

6 Queries

We now introduce the notion of a query to a rule extended knowledge base, that includes a DL knowledge base, a set of rules, and some facts.

Definition 1 *A* query *to a rule extended knowledge base is a (possibly ground) literal $q_{\mathbf{x}}$ with variables $\mathbf{x}$ (possibly empty). The* answer set *of $q_{\mathbf{x}}$ is the set of all substitutions of $\mathbf{x}$ with constants $\mathbf{c}$ from $\mathcal{A}$, such that the for each substitution the grounded query is logically implied by the rule extended knowledge base, i.e.,*

$$\{\mathbf{c} \text{ in } \mathcal{A} \mid \langle \Sigma, P \rangle \models q_{[\mathbf{x}/\mathbf{c}]}\}.$$

This definition of query is based on the notion of *certain answer* in the literature and it is very general. Given a Σ, we define *query rule* over Σ as a set of view rules together with a query literal selected from some head. In this way we capture the notion of a complex query expressed by means of a set of rules on top of an ontology.

The definition of query given above encompasses the different proposals of querying a DL knowledge base appeared in the literature. An important special case of query rule is with view acyclic DL axiom-based rules, which is better known as *conjunctive query* if each head literal appears only in one head, or *positive query* otherwise.

Recently, the Joint US/EU ad hoc Agent Markup Language Committee has proposed an OWL query language called OWL-QL [Fikes *et al.*, 2003], as a candidate standard language, which is a direct successor of the DAML Query Language (DQL). The query language is not fully formally specified, however it can be easily understood as allowing for conjunctive queries with distinguished variables (called

must-bind variables) and non distinguished variables (called *don't-bind* variables). In addition, *may-bind* variables apparently provide the notion of a *possible* answer as opposed to the *certain* answer which has been adopted in this paper. Query premises of OWL-QL allow to perform a simple form of local conditional query; this could be encoded as *assertions in DL queries*; see [Franconi and Tessaris, 2004].

7 Comparing the three approaches

We first show in this section the conditions under which the three approaches coincide. This corresponds essentially to the case of negation-free view rule-extended knowledge bases with empty TBoxes. Note that this is the case of pure Datalog without a background knowledge base, for which it is well known that the three different semantics give rise to the same answer set.

Theorem. *If we restrict a rule extended knowledge base with classical rules to view negation-free DL rules with TBox-free Σ, a rule extended knowledge base with lp-rules to NAF-free negation-free DL programs with TBox-free Σ, and a rule extended knowledge base with autoepistemic rules to view negation-free DL rules with TBox-free Σ, the semantics of the rule extended knowledge base with classical rules, with lp-rules, and with with autoepistemic rules coincide, i.e., the logical implication problem is equivalent in the three approaches.*

7.1 Examples

The above theorem is quite strict and it fails as soon as we release some assumption. We will show now by examples the differences between the three approaches. Consider the following knowledge base Σ, common to all the examples:

```
is-parent ≐ ∃is-parent-of
my-thing ≐ is-parent ⊔ ¬is-father
is-parent-of(john, mary)
is-parent(mary)
```

where we define, using standard DL notation, a TBox with the `is-parent` concept as anybody who is parent of at least some other person, and the concept `my-thing` as the union of `is-parent` and the negation of `is-father` (this should become equivalent to the top concept as soon as `is-father` becomes a subconcept of `is-parent`); and an ABox where we declare that John is a parent of Mary, and that Mary is parent of somebody.
Consider the following query rules, showing the effect of existentially quantified individuals coming from some TBox definition:

```
Q_ax(x) ← is-parent-of(x,y)
Q_lp(x) :- is-parent-of(x,y)
Q_ae(x) ⇐ is-parent-of(x,y)
```

The query `Q_ax(x)` returns {`john, mary`}; the query `Q_lp(x)` returns {`john`}; the query `Q_ae(x)` returns {`john, mary`}.
Consider now the query rules, which shows the impact of negation in the rules:

```
Q_ax(x,y) ← ¬is-parent-of(x,y)
Q_lp(x,y) :- ¬is-parent-of(x,y)
Q_ae(x,y) ⇐ ¬is-parent-of(x,y)
```

The query `Q_ax(mary, john)` returns false; the query `Q_lp(mary, john)` returns true; the query `Q_ae(mary, john)` returns false.
Consider now the following alternative sets of rules, which show that autoepistemic rules, unlike the axiom-based ones, do not influence TBox reasoning:

```
is-parent(x) ← is-father(x)
Q_ax(x) ← my-thing(x)

is-parent(x) ⇐ is-father(x)
Q_ae(x) ⇐ my-thing(x)
```

In the first axiom-based case, the query `Q_ax(paul)` returns true; in the second autoepistemic case the query `Q_ae(paul)` returns false (we assume that `paul` is an individual in Σ).

8 DL-based KBs and RDF

In [Franconi *et al.*, 2004] we recast the RDF model theory in a more classical logic framework, and use this characterisation to shed new light on the ontology languages layering in the semantic web. The ultimate purpose of this characterisation is to enable the integration of different rule and query semantics as presented in the previous sections with an RDF document base, in an attempt to solve the problem of interoperability between description logics based ontology languages (such as OWL-DL), rules/queries, and RDF. We have shown how the models of RDF can be related to the models of DL based ontology languages: this characterisation is fully compatible with the current semantics specification of both RDF (as defined in the RDF Model Theory (MT) in [Hayes, 2004]) and OWL-DL.

We first introduce the notion of minimal models for RDF graphs, and we use this notion to characterise RDF entailment: in fact, we prove that RDF entailment (as defined in the RDF MT) is equivalent to minimal models entailment. RDF minimal models can be associated to classical first order structures, that we call natural DL interpretations: these structures provide the semantic bridge between RDF and description logics based languages. The intuition beyond a natural DL interpretation is that it singles out the concepts and the individuals from an RDF minimal model – possibly in a polymorphic way, when the same URI is given both the meaning as a class and as an individual. For example, given the triple

⟨ex:o, rdf:type, ex:o⟩, its natural DL interpretation is such that the URI ex:o is interpreted as both a concept and an individual, and the individual ex:o is in the extension of the concept ex:o.

Once we have characterised RDF graphs in terms of their minimal models, it is possible to understand the notion of hybrid reasoning (e.g., logical implication or querying) with RDF graphs and DL knowledge bases. In particular, the answer of a query to an RDF graph given a DL-based ontology is defined as the standard DL-based ontology entailment restricted to the natural DL interpretations of the RDF graph. In [Franconi *et al.*, 2004] we prove an important reduction theorem: given an RDF graph $\mathcal{S}$ and a query Q, the answer set of Q to $\mathcal{S}$ as defined by RDF MT is the same as the intersection of the answers of Q to the (obvious) transformation of the natural DL interpretation of $\mathcal{S}$ into an ABox with the empty ontology. This shows a complete interoperability between RDF and DLs. For example, in absence of ontologies, it would be possible to use OWL-QL to answer queries to RDF graphs, or to use SPARQL to answer queries to ABoxes.

By exploiting the same technique the framework can be extended in order to accommodate the so called rule-extended knowledge bases discussed in the previous sections. Also in this case, the bridge is provided by the first order characterisation of RDF models.

References

[Fikes *et al.*, 2003] Richard Fikes, Patrick Hayes, and Ian Horrocks. OWL-QL - A Language for Deductive Query Answering on the Semantic Web. Technical report, Knowledge Systems Laboratory, Stanford University, Stanford, CA, KSL-03-14, 2003.

[Franconi and Tessaris, 2004] Enrico Franconi and Sergio Tessaris. Rules and queries with ontologies: a unified logical framework. In *Proc. of the Workshop on Principles and Practice of Semantic Web Reasoning (PPSWR'04)*, 2004.

[Franconi *et al.*, 2004] Enrico Franconi, Ian Horrocks, Birte Glimm, Lei Li, Jeff Z. Pan, Wolf Siberski, Giorgos Stamou, Sergio Tessaris, Vassilis Tzouvaras, and Holger Wache. D2.5.2: Report on query language design and standarization. Technical report, EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB, 2004. Available at http://knowledgeweb.semanticweb.org/.

[Grosof *et al.*, 2003] Benjamin N. Grosof, Ian Horrocks, Raphael Volz, and Stefan Decker. Description logic programs: Combining logic programs with description logic. In *Proc. of the Twelfth International World Wide Web Conference (WWW 2003)*, pages 48–57. ACM, 2003.

[Hayes, 2004] Patrick Hayes. RDF semantics. Technical report, W3C, February 2004. W3C recommendation, URL http://www.w3.org/TR/rdf-mt/.

[Horrocks and Patel-Schneider, 2004] Ian Horrocks and Peter F. Patel-Schneider. A proposal for an owl rules language. In *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*, 2004.