

Evaluating Datalog Tools for Meta-reasoning over OWL 2 QL

QURESHI HAYA MAJID and FABER WOLFGANG

University of Klagenfurt, Klagenfurt, Austria

(e-mails: hmq.17@hotmail.com, wolfgang.faber@aau.at)

submitted 14 March 2022; revised 8 November 2023; accepted 6 February 2024

Abstract

Metamodeling is a general approach to expressing knowledge about classes and properties in an ontology. It is a desirable modeling feature in multiple applications that simplifies the extension and reuse of ontologies. Nevertheless, allowing metamodeling without restrictions is problematic for several reasons, mainly due to undecidability issues. Practical languages, therefore, forbid classes to occur as instances of other classes or treat such occurrences as semantically different objects. Specifically, meta-querying in SPARQL under the Direct Semantic Entailment Regime uses the latter approach, thereby effectively not supporting meta-queries. However, several extensions enabling different metamodeling features have been proposed over the last decade. This paper deals with the Metamodeling Semantics (MS) over OWL 2 QL and the Metamodeling Semantic Entailment Regime (MSER), as proposed in Lenzerini *et al.* (2015, *Description Logics*) and Lenzerini *et al.* (2020, *Information Systems 88*, 101294), Cima *et al.* (2017, *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics*, 1–6). A reduction from OWL 2 QL to Datalog for meta-querying was proposed in Cima *et al.* (2017, *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics*, 1–6). In this paper, we experiment with various logic programming tools that support Datalog querying to determine their suitability as back-ends to MSER query answering. These tools stem from different logic programming paradigms (Prolog, pure Datalog, Answer Set Programming, Hybrid Knowledge Bases). Our work shows that the Datalog approach to MSER querying is practical also for sizeable ontologies with limited resources (time and memory). This paper significantly extends Qureshi and Faber (2021, *International Joint Conference on Rules and Reasoning*, Springer, 218–233.) by a more detailed experimental analysis and more background.

KEYWORDS: ontology, Datalog, metamodeling, rules

1 Introduction

The OWL ontology language (see for instance Horrocks *et al.* (2003); Antoniou and van Harmelen (2009)) provides a powerful data modeling language and automated reasoning abilities based on description logics (DL) (see for example Baader *et al.* (2003)). In the current version, OWL 2, several profiles exist that restrict the language features for limiting complexity of reasoning and easing implementations. In this article, we focus on the OWL 2 QL ontology language (see Motik *et al.* (2009)), which is geared toward

Ontology-Based Data Access (OBDA) Eiter and Krennwallner (2012), that is, querying a huge amount of preexisting data residing in possibly external information systems. OWL 2 QL is based on the Description Logic *DL-Lite_R* (see Calvanese *et al.* (2007)), and the tractability of main reasoning tasks with respect to the size of the data is guaranteed, see Motik *et al.* (2009). This is important, as the involved data can be very extensive in real-world applications.

An issue with ontologies that has gained momentum in recent years is meta-modeling and metaquerying, see for example Motik (2007), Brasileiro *et al.* (2016). Metamodeling refers to higher level abstractions in ontologies, where classes, called metaclasses, can have other intensional predicates, that is, classes, properties or datatypes, as instances, and properties can have as instances pairs including intensional predicates as well. Metaqueries are queries that range over intensional predicates and not only over individuals.

The language OWL 2, and therefore also OWL 2 QL, syntactically allows for meta-modeling employing punning.¹ With punning, the same name can be used for ontology elements of different types (most notably, classes and individuals). However, the standard semantics of OWL 2, the Direct Semantics (DS) Horrocks *et al.* (2012), treats punning in a way that is not intended by meta-modeling. The reason is that DS imposes a sharp distinction between the entity types, interpreting the differently typed occurrences of the same name necessarily as different entities.

The following example, adapted from Guizzardi *et al.* (2015), shows the use of meta-modeling in OWL 2 QL and how it affects query answering under DS.

Example 1

Consider the modeling of biological species, stating that all GoldenEagles are Eagles, all Eagles are Birds, and that Harry is an instance of GoldenEagle, which by inference is also an instance of Eagle and Birds.

```
SubClassOf(:Eagle :Birds)
SubClassOf(:GoldenEagle :Eagle)
ClassAssertion(:GoldenEagle :Harry)
```

However, in the species domain one can also articulate expressions about species themselves, such as “GoldenEagle is listed in the IUCN Red List of endangered species”. This states that GoldenEagle as a whole class is an endangered species. This is different from specific properties of species and also does not state anything about individuals, such as Harry. To formally model this statement, we can declare GoldenEagle to be an instance of the new class EndangeredSpecies.

```
ClassAssertion(:EndangeredSpecies :GoldenEagle)
```

Asserting GoldenEagle to be of type EndangeredSpecies is concise and clear. Note that making GoldenEagle a subclass of EndangeredSpecies would not be correct, as it results in incorrect conclusions like “Harry is an EndangeredSpecies”. In the assertion above, EndangeredSpecies is a metaconcept for GoldenEagle.

Despite the syntactic support of punning offered by OWL 2 QL, DS prevents extracting some information from the knowledge base by means of metaquerying. For instance,

¹ <http://www.w3.org/2007/OWL/wiki/Punning>.

querying “Is Harry an individual of a species that is endangered?” over the ontology of Example 1 is not really possible.

Let us consider this in a bit more detail. SPARQL (see, for example Pérez *et al.* (2009)) is the de-facto standard ontology query language that supports basic graph patterns (BGP). Simple conjunctive queries are allowed in the language; however, there are type restrictions to be met in DS. Suppose that one formulates a metaquery, for simplicity denoted as a conjunctive query in the style of relational calculus:

$$\{(z) \mid \exists y ES(y) \wedge y(z) \wedge Lives_in(z, CPZ)\}$$

The intention is to retrieve all individuals that live in Central Park Zoo (CPZ) and are instances of a species that is endangered. The corresponding SPARQL query would be:

```
SELECT ?z WHERE {
    ?y a ES.
    ?z a ?y.
    ?z Lives_in CPZ
}
```

Unfortunately, this query is not allowed under the *Direct Semantics Entailment Regime* (DSE_R) Glimm (2011), which is the logical underpinning for SPARQL queries for interpreting the logical theories corresponding to OWL 2 ontologies under DS. DSE_R imposes the so-called *Variable Typing*² constraint that disallows the use of the same variable in a different type of position and makes metaquerying impossible. In the relational calculus notation above it becomes clear that this query has a higher order aspect to it, as a variable occurs in the position of a predicate.

To remedy the limitation of metamodeling, higher order semantics (HOS) was introduced in Lenzerini *et al.* (2015) for OWL 2 QL ontologies. The interpretation structure of HOS follows the Hilog-style semantics of Chen *et al.* (1993), which allows the elements in the domain to have polymorphic characteristics. Data complexity of HOS stays in AC⁰ for answering unions of conjunctive queries, as shown in Lenzerini *et al.* (2016b). In Lenzerini *et al.* (2020), HOS is referred to as Meta-modeling Semantics (MS), which is the terminology that we will adopt in this paper. And to remedy the limitation of metaquerying, the Meta-modeling Semantics Entailment Regime (MSE_R) was proposed in Cima *et al.* (2017), which does allow meta-modeling and meta-querying using SPARQL over OWL 2 QL.

In Cima *et al.* (2017), a reduction from query-answering over OWL 2 QL to Datalog queries is provided, and experimental results using two Datalog engines, LogicBlox, see Aref *et al.* (2015), and RDFox, see Nenov *et al.* (2015), are reported. Datalog (see for instance Ceri *et al.* (2012)) is an attractive target language, as it has received increased academic interest with renewed tool support over the last few decades, see for instance Huang *et al.* (2011), De Moor *et al.* (2012). From logic programming (cf. Lloyd (2012)), Datalog emerged as an eminent language. Datalog enjoys tractable reasoning with respect to the input data size (given a fixed program). The key aspect of Datalog is to be utilized as an expressive querying language for relational data. As a result, Datalog has advanced into a first-class formalism with efficient implementations. Moreover, Datalog has also

² <https://www.w3.org/TR/2013/REC-sparql11-entailment-20130321/>.

proven particularly relevant for Semantic Web applications such as ontological modeling and reasoning, as rules in Datalog can represent clauses in the fragment of function-free Horn First-Order Logic.

In this work, we address the practical effectiveness of the Datalog reduction of [Cima et al. \(2017\)](#). In [Cima et al. \(2017\)](#), a preliminary performance analysis was reported, using the tools LogicBlox and RDFox on queries over LUBM ontologies. In [Lenzerini et al. \(2020\)](#), meta-modeling and meta-querying in OWL 2 QL under MSER was evaluated as well using the larger MODEUS ontologies, which also contain more meta-modeling. However, that work did not evaluate the Datalog characterization of [Cima et al. \(2017\)](#). In our work, we consider on the one hand a larger spectrum of tools that support Datalog (but stem from a variety of foundations and techniques), and on the other hand consider queries over both LUBM and MODEUS ontologies, expanding [Qureshi and Faber \(2021\)](#). It turns out that other tools perform very well, too, especially in the meta-modeling heavy MODEUS scenarios, which were not considered in [Cima et al. \(2017\)](#). So there are tools that allow for practical query answering also on sizable ontologies.

Contributions. This work builds upon [Cima et al. \(2017\)](#) and [Qureshi and Faber \(2021\)](#) and extends these by a detailed study of metamodeling and metaquerying in OWL 2 QL, along with the experiments and results. The main contributions of this work can be summarised as follows:

- We have extended the performance evaluation of the Datalog-based MSER approach discussed in [Qureshi and Faber \(2021\)](#) to three more back-end tools, the hybrid reasoning tools DLVHex and HexLite, and the lazy-grounding ASP solver Alpha, totalling nine back-end tools.
- In [Qureshi and Faber \(2021\)](#), no time and memory limit was set. In this work, we consider two different limits, a *Tight Resource Limit* and a *More Generous Resource Limit*.
- In the evaluation, we considered both meta-modeling ontologies and meta-queries. Furthermore, we have also considered a non-metamodeling ontology and standard query setting. The latter serves as a baseline, in order to understand the feasibility of MSER also in a settings in which it is equivalent to DSER.
- We give a detailed discussion on the performance and the evaluation results of the tools under the two limits.

Outline. The remainder of this paper is organized as follows. Section 2 introduces the background notions like OWL 2 QL, MS, SPARQL, and Datalog. Section 3 gives an overview of the rewriting presented in [Cima et al. \(2017\)](#), [Qureshi and Faber \(2021\)](#). Section 4 illustrates the series of experiments we have carried out using the LUBM and MODEUS ontologies and presents a detailed discussion of the findings. Section 5 compares our approach with other approaches to metamodeling in the literature. Section 6 concludes the paper with a brief discussion and pointing out some future work.

2 Preliminaries

In this section, we recall the preliminary notions that are needed for the rest of the paper, in particular OWL 2 QL under MS, SPARQL, and Datalog.

Table 1. OWL 2 QL axioms

TBox (positive axioms)	TBox (negative axioms)
$\mathbf{b} \sqsubseteq_c \mathbf{c}$ (class inclusion)	$\mathbf{b}_1 \sqsubseteq_c \neg \mathbf{b}_2$ (class disjointness)
$\mathbf{p}_1 \sqsubseteq_p \mathbf{p}_2$ (object property inclusion)	$\mathbf{p}_1 \sqsubseteq_p \neg \mathbf{p}_2$ (object property disjointness)
$\mathbf{ref}(\mathbf{p})$ (reflexive object property)	$\mathbf{irref}(\mathbf{p})$ (irreflexive object property)
	ABox axioms
$\mathbf{ca}(\mathbf{i})$ (class assertion)	$\mathbf{ra}(\mathbf{i}_1, \mathbf{i}_2)$ (object property assertion)

2.1 OWL 2 QL

This section reviews the syntax of the ontology language OWL 2 QL and the *Meta-modeling Semantics* (MS) for OWL 2 QL, a profile of OWL 2. In general, OWL 2 QL is an easy-to-implement sub-language of OWL 2, obtained by decreasing the expressive power of OWL 2 DL, matching *DL-Lite_R*, allowing for tractable reasoning and querying under direct semantics in AC⁰ (data complexity). In this work we did not deal with data properties and literals and that is why did not include them in language definitions. We refer to [Lenzerini et al. \(2021\)](#) for a complete language definition.

Syntax. We start by recalling some basic elements used for representing knowledge in ontologies: *individuals* (i), objects of a domain of discourse, *concepts* or *classes* (c), sets of individuals with common properties, and *roles* or *object properties* (p), relations that link individuals. Given the vocabulary V of an OWL 2 QL ontology $V = (V_n, V_c, V_p, V_i)$, where V_n consists of Internationalized Resource Identifiers (IRIs), which are used for denoting classes, object properties, and individuals, grouped into sets V_c, V_p, V_i , respectively. V_c (resp., V_p) includes the special symbols \top_c (resp., \perp_c) and \top_p (resp., \perp_p). We denote $Exp(V)$ the set of all well-formed expressions over V written as $Exp(V) = V_n \cup Exp_c(V) \cup Exp_p(V) \cup V_i$, where

- $Exp_c(V) = V_c \cup \{\exists e_1.e_2 \mid e_1 \in Exp_p(V), e_2 \in V_c\}$ are class expressions.
- $Exp_p(V) = V_p \cup \{e^- \mid e \in V_p\}$ are object property expressions, containing either object properties or their inverses.
- the set of individual expressions coincides with V_i .

An OWL 2 QL *ontology* is a finite set of logical axioms over V . The axioms allowed in an OWL 2 QL ontology have one of the form listed in the Table 1, where (with optional subscripts) $\mathbf{ca}, \mathbf{ra}, \mathbf{i}$ denote an atomic class, an atomic object property, an individual expression; \mathbf{b} denotes a class expression; \mathbf{c} and \mathbf{p} denote a class and an object property. Moreover, \mathbf{ref} and \mathbf{irref} are keywords for declaring object properties to be reflexive and irreflexive, respectively. Table 1 is divided into TBox axioms (further divided into positive TBox axioms and negative TBox axioms) and ABox axioms.

In Table 1, (i) the positive TBox axioms are the inclusion axioms, namely class inclusions, object property inclusions, plus the reflexivity declarations, (ii) negative TBox axioms are the disjoint axioms, namely class disjointness, object property disjointness, plus irreflexivity declarations, and (iii) ABox axioms are membership assertion axioms, namely class membership assertion, object property membership assertion and different individual declarations. For simplicity, in this paper, we omit to deal with OWL 2

QL axioms that can be expressed by appropriate combinations of the axioms specified in Table 1. Examples of such axioms are: equivalences, disjointness of more than two entities, or object property symmetry. Lastly, we assume that the symbols ca , ra appearing in the ABox axioms of the ontology such as $ca(i)$ and $ra(i_1, i_2)$, also appear in the TBox of the same ontology without loss of generality (by having axioms like $ca \sqsubseteq_c \top_c, ra \sqsubseteq_p \top_p$).

Now, we return to the *punning* feature (the same entity can appear on the position of different types) in OWL 2 QL discussed in the Introduction. Every position in the axioms refers to a certain type, for example $c_1 \sqsubseteq_c c_2$ is the *SubClassOf* axiom, where both positions are occupied by class expressions, therefore both are *class* positions. Similarly $ca(i)$ is a *ClassAssertion* axiom where first position in the axiom is a *class* position and the second one is an *individual* position. If we look at Example 1, the entity GoldenEagle appears as the first argument of a *SubClassOf* axiom and the second argument of a *ClassAssertion* axiom, thus the former position refers to a *class* position and the latter refers to an *individual* position.

2.2 Meta-modeling semantics

The MS is based on the idea that every entity in V may simultaneously have more than one type, so it can be a class, or an individual, or data property, or an object property or a data type. To formalize this idea, the MS has been defined for OWL 2 QL. We refer to Lenzerini *et al.* (2021) for a comprehensive definition.

The MS for \mathcal{O} over V is based on the notion of interpretation, constituted by a tuple $\mathcal{I} = \langle \Delta, \cdot^I, \cdot^C, \cdot^P, \cdot^T \rangle$, where $(\mathcal{P}(S))$ denotes the power set of S

- Δ is the union of the two nonempty disjoint sets: $\Delta = \Delta^o \cup \Delta^v$, where Δ^o is the object domain, and Δ^v is the value domain. Each element in the world is either object or a value;
- $\cdot^I : \Delta^o \rightarrow \{True, False\}$ is a total function for each object $o \in \Delta^o$, which indicates whether o is an individual; if $\cdot^C, \cdot^P, \cdot^D, \cdot^T$ are undefined for some o , then we require $o^I = True$, also in a few other cases such as if o is in the range of \cdot^C ;
- $\cdot^C : \Delta^o \rightarrow \mathcal{P}(\Delta^o)$ is partial and can assign the extension of a class;
- $\cdot^P : \Delta^o \rightarrow \mathcal{P}(\Delta^o \times \Delta^o)$ is partial and can assign the extension of an object property;
- \cdot^T is a function that maps every expression in $Exp(V)$ to Δ^o and every literal to Δ^v .

There are a couple of constraints on interpretations, for which we refer to Table 3 in Lenzerini *et al.* (2021). Note that we follow Lenzerini *et al.* (2021) also with naming the interpretation \mathcal{I} while using the same symbol for one of its component functions, which can be distinguished by the fact that the latter is used only in superscripts.

This allows for a single object o to be simultaneously interpreted as an individual via \cdot^I , a class via \cdot^C and an object property via \cdot^P . For instance, for Example 1, \cdot^C, \cdot^I would be defined for *GoldenEagle* ^{\mathcal{I}} , while \cdot^P would be undefined for it.

The semantics of a logical axiom α is defined in accordance with the notion of axiom satisfaction for the MS interpretation \mathcal{I} . The language is specified in Table 3.B in Lenzerini *et al.* (2021). \mathcal{I} is a model of ontology \mathcal{O} if it satisfies all axioms of \mathcal{O} . Finally, an axiom α is said to be logically implied by \mathcal{O} , denoted as $\mathcal{O} \models \alpha$, if it is satisfied by every model of \mathcal{O} .

2.3 The query language

As query language, we consider conjunctive queries expressed using SPARQL. Conjunctive queries are based on the notion of query atoms, which is based on the notion of query terms (may involve variables). Given an ontology over V and set of variables \mathcal{V} disjoint from V , a query atom (or atom) has the same form as axioms with the difference that query terms (expressions of the set $Exp(V \cup \mathcal{V})$) may occur in it.

A conjunctive SPARQL query q is of the form

$$SELECT ?v_1 \dots ?v_n \text{ WHERE } \{B\}$$

where $\{?v_1, \dots, ?v_n\} \subseteq \mathcal{V}$ and B called body of conjunctive query q , denoted as $body(q)$, consisting of a conjunction of triples that are to be read as atoms over terms $Exp(V \cup \mathcal{V})$, and $\{?v_1, \dots, ?v_n\}$ has to be a subset of the variables occurring in B . Given an ontology \mathcal{O} and a conjunctive SPARQL query q , the answers to q are the tuples $\langle ?v_1\sigma, \dots, ?v_n\sigma \rangle$, where σ is a substitution that maps all variables in q to elements of V such that $\mathcal{O} \models B\sigma$.

Note that SPARQL usually uses RDF triple syntax, which we will also write as atoms. In the following we will sometimes use the notation $q(v_1, \dots, v_n) \leftarrow B'$ to denote a query of the form above, where B' corresponds to B but is written as a conjunction of atoms.

We would like to distinguish between standard queries and meta-queries. A meta-query is a query consisting of meta-predicates p and meta-variables v , where p can have other predicates as their arguments and v can appear in predicate positions. For instance, if c is a class name and x, y are variables, query

$$q(x) \leftarrow c(x) \wedge x(y).$$

will return all classes that are themselves members of another class.

The simplest form of meta-query is a query where variables appear in class or property positions also known as *second-order queries*. More interesting forms of meta-queries allow one to extract complex patterns from the ontology, by allowing variables to appear simultaneously in individual object and class or property positions. We will refer to non-meta-queries as standard queries.

As for the semantics of query \mathcal{Q} under MS, we rely on the definitions of [Lenzerini et al. \(2021\)](#) and do not provide any formal details here. Essentially, given a query \mathcal{Q} expressed over \mathcal{O} , $\mathcal{O} \models_{MS} \mathcal{Q}\sigma$ holds if $\mathcal{Q}\sigma$ is true in every MS model of \mathcal{O} for a substitution σ for the variables in \mathcal{Q} .

2.4 Datalog

Datalog is a declarative query language rooted in logic programming. Datalog has similar expressive power as SQL, whose more recent versions also allow to express recursive queries over a relational database, previously a distinguishing feature of Datalog.

Syntax.

A Datalog program consists of a finite set of facts and rules, where facts are seen as the assertions about the world, that is, “*Harry is an instance of GoldenEagle*” and rules are sentences that allow for deducing facts from other facts, that is, “*If X is an instance-Of*

Y , and Y has a parent Z , then X is an instance-Of Z ". Generally, rules contain variables like X , Y , or Z , and when combining both facts and rules, they form a knowledge base.

A Datalog program is a finite set of rules r , where both facts and rules are represented as Horn clauses of the form

$$h \leftarrow b_1, \dots, b_n.$$

In r , h is the rule head and b_1, \dots, b_n is the rule body. Each of b_i and h are atoms of the form $p(t_1, \dots, t_n)$, where p is a predicate symbol and t_i are terms that can be constants or variables; n is the arity of p . We say that r is a fact if it has an empty body and omit the symbol \leftarrow in that case, for example

$$\text{instc}(\text{GoldenEagle}, \text{Harry}).$$

The body can be thought of as a conjunction, for example in

$$\text{instc}(Z, X) \leftarrow \text{instc}(X, Y), \text{isacCC}(Y, Z).$$

In these examples, *GoldenEagle*, *Harry* are the constants, X , Y , Z are the variables and *instc*, *isacCC* are the predicates. A literal, fact, or rule, which does not contain any variables, is called ground.

Rules need to satisfy safety in order to be domain-independent and to guarantee that the grounding of the Datalog program is finite, so each variable occurring in the rule head must also occur in the rule body and facts may not contain variables.

Semantics

For a Datalog program P , the Herbrand Base $HB(P)$ is the set of all atoms of the form $p(c_1, \dots, c_n)$, where p and c_i are predicates and constants in P . An interpretation $I \subseteq HB(P)$ consists of those atoms that are true in I . A positive ground literal l (resp. $\neg l$) is satisfied by I if $l \in I$ (resp. $l \notin I$). A rule r is satisfied if $h \in I$ whenever $\{b_1, \dots, b_n\} \subseteq I$, and an interpretation I is a model of a program P if all of its rules are satisfied. The semantics of P is given by the subset-minimal model of P , which is guaranteed to exist and to be unique.

A list of ground atoms a_1, \dots, a_n , interpreted as conjunction of atoms are said to be logical consequences of P denoted as $P \models a_1, \dots, a_n$ if each a_i is satisfied in all models of P , which means that each a_i is satisfied in the subset-minimal model.

A query for a Datalog program D is a conjunction of atoms written as

$$q_1, \dots, q_n?$$

It consists of atoms as in Datalog rules, and the answers with respect to a program P are those substitutions σ (mapping variables in the query to constants) such that $\{q_1\sigma, \dots, q_n\sigma\} \subseteq M$ for the subset-minimal model M of P . Equivalently, the answers are those σ such that $P \models q_1\sigma, \dots, q_n\sigma$.

While there are dedicated Datalog-only systems, Datalog is also the core of several richer languages and systems that support them. This is the main motivation of our work, in which we examine how various tools that support Datalog query answering perform on Datalog queries stemming from meta-querying over OWL 2 QL ontologies.

Table 2. τ function

$\mathcal{P}_{\mathcal{O}}^{ql}$	α	$\tau(\alpha)$	α	$\tau(\alpha)$
$\mathcal{P}_{\mathcal{O}}^{ql,\mathcal{T}}$	$c1 \sqsubseteq c2$	isacCC(c1, c2)	$r1 \sqsubseteq \neg r2$	disjrRR(r1,r2)
	$c1 \sqsubseteq \exists r2^-.c2$	isacCI(c1, r2, c2)	$c1 \sqsubseteq \neg c2$	disjcCC(c1,c2)
	$\exists r1 \sqsubseteq \exists r2.c2$	isacRR(r1,r2,c2)	$c1 \sqsubseteq \neg \exists r2^-$	disjcCI(c1,r2)
	$\exists r1^- \sqsubseteq c2$	isacIC(r1,c2)	$\exists r1 \sqsubseteq \neg c2$	disjcRC(r1,c2)
	$\exists r1^- \sqsubseteq \exists r2.c2$	isacIR(r1,r2,c2)	$\exists r1 \sqsubseteq \neg \exists r2$	disjcRR(r1,r2)
	$\exists r1^- \sqsubseteq \exists r2^-.c2$	isacII(r1,r2,c2)	$\exists r1 \sqsubseteq \neg \exists r2^-$	disjcRI(r1,r2)
	$r1 \sqsubseteq r2$	isarRR(r1,r2)	$\exists r1^- \sqsubseteq \neg c2$	disjcIC(r1,c2)
	$r1 \sqsubseteq r2^-$	isarRI(r1,r2)	$\exists r1^- \sqsubseteq \neg \exists r2$	disjcIR(r1,r2)
	$c1 \sqsubseteq \exists r2.c2$	isacCR(c1,r2,c2)	$\exists r1^- \sqsubseteq \neg \exists r2^-$	disjcII(r1,r2)
	$\exists r1 \sqsubseteq c2$	isacRC(r1,c2)	$r1 \sqsubseteq \neg r2^-$	disjrRI(r1,r2)
	$\exists r1 \sqsubseteq \exists r2^-.c2$	isacRI(r1,r2,c2)	irref(r)	irrefl(r)
		refl(r)		
$\mathcal{P}_{\mathcal{O}}^{ql,\mathcal{A}}$	$c(x)$	instc(c,x)	$x \neq y$	diff(x,y)
	$r(x, y)$	instr(r,x,y)		

3 Overview of the approach

In this section, we recall query answering under the Meta-modeling Semantics Entailment Regime (MSER) from [Cima et al. \(2017\)](#). This technique reduces SPARQL query answering over OWL 2 QL ontologies to Datalog query answering. The main idea of this approach is to define (i) a translation function τ mapping OWL 2 QL axioms to Datalog facts and (ii) a fixed rule base \mathcal{R}^{ql} that captures inferences in OWL 2 QL reasoning. Note that, for the sake of simplicity, we follow [Cima et al. \(2017\)](#) and do not consider ontologies containing data properties for easier presentation.

The reduction employs a number of predicates, which are used to encode the basic axioms available in OWL 2 QL. This includes both axioms that are explicitly represented in the ontology (these will be added to the Datalog program as facts resulting from the mapping τ) and axioms that logically follow (these will be derivable by the fixed rules \mathcal{R}^{ql}). In a sense, this representation is closer to a meta-programming representation than other Datalog embeddings that translate each axiom to a rule.

The function τ is used to encode the OWL 2 QL assertions α as facts. For a given ontology \mathcal{O} , we will denote the set of facts obtained by applying τ to all of its axiom as $\mathcal{P}_{\mathcal{O}}^{ql}$; it will be composed of two portions $\mathcal{P}_{\mathcal{O}}^{ql,\mathcal{T}}$ and $\mathcal{P}_{\mathcal{O}}^{ql,\mathcal{A}}$, as indicated in Table 2.

The fixed Datalog program \mathcal{R}^{ql} can be viewed as an encoding of axiom saturation in OWL 2 QL. The full set of rules were provided by the authors of [Cima et al. \(2017\)](#) and reported in [Qureshi and Faber \(2021\)](#). They can also be found in the online repository <https://doi.org/10.5281/zenodo.7286886>. We will consider one rule to illustrate the underlying ideas:

$$\text{isacCR}(c1,r2,c2) \text{ :- isacCC}(c1,c3), \text{isacCR}(c3,r2,c2)$$

The above rule encodes the following inference rule:

$$\mathcal{O} \models c1 \sqsubseteq c3, \mathcal{O} \models c3 \sqsubseteq \exists r2.c2 \Rightarrow \mathcal{O} \models c1 \sqsubseteq \exists r2.c2$$

Finally, the translation can be extended in order to transform conjunctive SPARQL queries under MS over OWL 2 QL ontologies into a Datalog query. For example, consider the following query that retrieves all triples $\langle x, y, z \rangle$, where x is a member of class y that is a subclass of z :

```
SELECT ?x ?y ?z WHERE {
  ?x rdf:type ?y.
  ?y rdfs:SubClassOf ?z
}
```

This can be translated to a Datalog query

$$\text{instc}(X,Y), \text{isacCC}(Y,Z)?$$

In general, these queries will be translated into a rule plus an atomic query to account for projections. The previous example will be translated to

$$q(X,Y,Z) \leftarrow \text{instc}(X,Y), \text{isacCC}(Y,Z). \\ q(X,Y,Z)?$$

4 Experimental setup

In this section, we describe the extended experiments that we have conducted, including the tools we used, the ontologies and queries we considered, and report on the outcomes. All material is available at <https://doi.org/10.5281/zenodo.7286886>.

4.1 Tools

We have implemented the translation of ontology axioms summarized in Table 2 in Java. We should point out that this implementation is not optimized and serves as a proof of concept. For the Datalog back-end, we have tested six tools from Qureshi and Faber (2021) and add three additional tools in our experiments. Our decisions with respect to tool selection were based on problem statement analysis and preliminary performance assessments. These tools stem from different logic paradigms like Prolog, pure Datalog, Answer Set Programming (ASP), and Hybrid-Knowledge Bases. In the following, we briefly describe each of these tools. For more detailed descriptions, we refer interested readers to Qureshi and Faber (2021).

4.2 Datalog

A common practice across all these Datalog systems is the use of the Datalog language in the area of efficient query answering, formal reasoning, and analysis, where tools such as RDFox and LogicBlox have shown great results.

4.2.1 RDFox

RDFox Nenov *et al.* (2015) is an in-memory, scalable, centralized data engine for Resource Description Framework (RDF) data models. The tool supports the current standard

querying language SPARQL 1.1. It also allows for reasoning and representing knowledge in rules, supporting materialization-based parallel Datalog reasoning. RDFox uses parallel reasoning algorithms to support Datalog reasoning over RDF data.

4.2.2 LogicBlox

LogicBlox Aref *et al.* (2015) is another state-of-the-art Datalog engine to allow applications to automate and enhance their decision making via a single expressive declarative language. LogicBlox supports module mechanisms and uses non-trivial type systems to detect programming errors and aid query optimization de Moor *et al.* (2008).

4.3 Prolog

Prolog is the classic logic programming language. Datalog was motivated by Prolog and can be viewed as a subset of it. While using Prolog naively on Datalog queries will lead to termination issues, techniques such as tabling alleviate this. Its application in various problem-solving areas has demonstrated its capabilities due to increasingly efficient implementations and ever richer environments such as XSB.

4.3.1 XSB

XSB Swift and Warren (2012) is a logic programming engine and language rooted in Prolog. It supports Prolog's standard functionality, and features a powerful technique called tabling, which significantly increases its applicability and is particularly relevant for Datalog query answering. Since it relies on a top-down technique, its internals are significantly different from RDFox and LogicBlox.

4.4 Hybrid approaches

Hybrid approaches are interesting because they support the DL-based knowledge representation languages at the core of the semantic web, and at the same time, they overcome some shortcomings of DLs, where rules are better suited. The hybrid solutions range from *loosely coupled* formalisms, such as those in Eiter *et al.* (2008), Wang *et al.* (2004), where the rule parts and DL parts of the knowledge base are treated like separate components that interact by querying each other, to fully integrated approaches, or *tightly coupled* formalisms such as Bruijn *et al.* (2011), Motik and Rosati (2007), where the vocabulary and semantics of rule and DL parts are homogeneous. In this work, we focus on the hybrid formalisms NoHR.

4.4.1 NoHR

NoHR Lopes *et al.* (2017), the Nova Hybrid Reasoner allows to query a combination of DL ontologies (OWL 2 EL or OWL 2 QL) and non-monotonic rules in a top down fashion. NoHR is the first hybrid reasoner of its kind for Protégé. We have included NoHR because we wanted to assess the overhead it produces with respect to XSB, as the agenda for our future work includes leveraging hybrid tools like NoHR for meta-modeling and meta-querying.

4.5 Answer Set Programming

ASP Brewka *et al.* (2011) is a declarative programming paradigm that applies non-monotonic reasoning. ASP falls into the category of declarative and logic programming and can be viewed as an extension to Datalog, which handles features such as negation as failure, disjunction in a semantically clear way. Its main applications have been in solving combinatorial problems, but it has also been successfully used for reasoning in knowledge representation and databases. The success of ASP is partially due to efficient solver technology available for evaluation, for instance Clingo, DLV2, Alpha, DLVHex, and HexLite. Most state-of-the-art ASP systems follow the *ground and solve* approach, where an input program is turned into a corresponding variable-free (ground) program for which answer sets are then computed. A few ASP systems use a *lazy-grounding* technique Weinzierl *et al.* (2019); Dao-Tran *et al.* (2012), where the main idea is to interleave grounding with solving and generate only ground rules necessary in each position of the search space, for instance *Alpha*.

4.5.1 Clingo

Clingo Gebser *et al.* (2010) is probably the currently most widely used tool for ASP. It follows the ground and solve approach, with subsystems *gringo* and *clasp* for grounding and solving, respectively.

4.5.2 DLV2

DLV2 Alviano *et al.* (2017) is another ASP system with particular emphasis on Disjunctive Logic Programming and based on the declarative programming language datalog, known for being a convenient tool for knowledge representation. It is also based on the ground and solve approach.

4.5.3 Alpha

Alpha Weinzierl (2017) is an ASP system that reads logic programs and computes the corresponding answer sets. In contrast to many other ASP systems, *Alpha* implements a lazy-grounding approach to overcome memory limits when working with considerable large input. Lazy grounding is a novel way to address the bottleneck in ground and solve approaches, by attempting to ground only what is needed. *Alpha* is not the quickest framework available, as it is intended as a testbed and prototype for lazy-grounding techniques. However, there are domains in which *Alpha* succeeds, where ground and solve systems fail completely.

4.5.4 DLVHex

DLVHex Eiter *et al.* (2006), a logic programming based reasoner, computes the models of so-called HEX-programs. The HEX-programs extend the ASP paradigm using non-monotonic logic programs with bidirectional access to external sources. Intuitively, the logic program sends information given by constants or predicate extensions to the external source, giving back output values imported into the program. It uses the *conflict-driven*

technique for the evaluation of a program, to find an assignment that satisfies all *nogoods* (set of literals which must not be true simultaneously, see Gebser *et al.* (2012)). The *DLVHex* system in its current version supports all features defined in the ASP standard, including function symbols, choice rules, conditional literals, aggregates, and weak constraints.

4.5.5 *HexLite*

The *HexLite* Schüller (2019) solver is a lightweight alternative to *DLVHex* with a restricted set of external computations. The intention is to provide a lightweight system for an easy start with HEX. The main aim of *HexLite* is to achieve efficiency and simplicity, both in the implementation and installation of the system. It uses the *pragmatic-hex-fragment* that permits to separate external computations into two parts: first, deal with those that can be evaluated during the instantiation, and second, deal with those that can be evaluated during the search. *HexLite* is implemented in Python and uses the *Clingo* Python API as a back-end for ASP grounding and search.

4.6 Computational settings

The versions of the evaluated systems are: RDFox 5.1.0, LogicBlox 4.34.0, XSB 4.0, NoHR 3.0, Clingo 5.2.2, DLV2 2.0, Alpha 0.5.0, DLVHex 2.5.0, HexLite 1.4.1 (and the hexlite-owl-api plugin version is 1.2).

We observe that the tools slightly differ in the Datalog syntax they require and thus needed minor adjustments. For instance, LogicBlox uses `<-` instead of `:-`, and also variables are denoted in different ways in the various input languages. RDFox and XSB needed some major adjustments, which are described in detail in Qureshi and Faber (2021). All other tools received the same input, except for variations due to language syntax.

4.6.1 Benchmark

Instead of creating a new benchmark from scratch, we decided to build our experiments on top of two datasets.

Lehigh University Benchmark (LUBM) was developed to facilitate the evaluation of Semantic Web reasoners in a standard and systematic way, with respect to extensional queries over synthetic OWL data scalable to arbitrary sizes, see Guo *et al.* (2005). It describes a university domain with departments, courses, students, and faculty information. Although LUBM is a fairly simple ontology and does not contain meta-axioms, it captures existentially quantified knowledge that we find interesting to evaluate with different tools under MSER. Please note that MSER is not restricted to metamodeling features in ontology only, and we used LUBM to show the performance of different tools with non-metamodeling ontology. LUBM comes with a predefined data generator to generate random sizes of \mathcal{A} , which can be used to test the system ability of handling data of varying sizes. We have used LUBM(1) consisting of 43 classes, 32 properties (including 25 object properties and 7 datatype properties) and 10,334 axioms and LUBM(9) consisting of 43 classes, 32 properties (including 25 object properties and 7 datatype properties)

and 79,501 axioms. The main difference between LUBM(1) and LUBM(9) is the size of the ABox. It uses OWL Lite language features including *inverseOf*, *someValuesFrom* restrictions, and *intersectionOf*.

Making Open Data Effectively USable (MODEUS) contains four ontologies describing the *Italian Public Debt* domain, designed in a joint project with the Italian Ministry of Economy and Finance, with information like financial liability or financial assets to any given contracts, see [Lenzerini et al. \(2020\)](#). The MODEUS ontologies are relatively new and more complex than LUBM as it contains not only the meta-axioms but also complex patterns equipped with a spectrum of metamodeling features. We want to clarify that the MODEUS ontologies are not standard benchmark ontologies; we are using these ontologies as a benchmark for this work. The base ontology was progressively extended to obtain three larger ontologies. The first ontology MEF_00 contains 92 classes, 40 meta-classes, 11 properties, about 27,000 individuals, 479 TBox axioms and about 13,7000 axioms. The second ontology MEF_01 obtained from MEF_00 by adding only 7000 fresh individuals and 35,000 ABox axioms, which makes about 33,000 individuals, about 172,000 ABox axioms and the rest is same in that it contains 92 classes, 40 meta-classes and 11 properties. The third ontology MEF_02 obtained from MEF_00 by adding new classes, subset classes and disjoints, which makes 97 classes, 43 meta-classes, 11 properties, about 26,000 individuals, 542 TBox axioms, and about 137,000 ABox axioms. The fourth ontology MEF_03 obtained from MEF_02 by adding 4000 individuals and 20,000 ABox axioms, which give us 98 classes, 43 meta-classes, 11 properties, about 30,000 individuals, 542 TBox axioms, and about 157,000 ABox axioms. MODEUS ontologies are meta-modeling ontologies with meta-classes and meta-properties.

4.7 Benchmark queries

For the queries, we used the queries given by these datasets and some additional queries (or meta-queries) for LUBM. We referred the queries to their proper sites.

The LUBM dataset comes with standard 14 queries³ that contain queries with variables only in place of individuals. These queries represent a variety of properties and several performance metrics like low selectivity vs high selectivity, implicit relationships vs explicit relationships, small input vs large input, etc. We have also considered the queries *mq1*, *mq4*, *mq5*, and *mq10* from [Kontchakov et al. \(2014\)](#) and called them meta-queries as they comprise variables in predicate position that is over class and property names. We have also considered two special case queries and called them special case meta-queries⁴ *sq1*, and *sq2* from [Cima et al. \(2017\)](#) to test the meta-modeling features and identify the new challenges introduced by the additional expressivity over the ABox queries.

In special-case meta-queries, we check the impact of meta-classes and DISJOINTWITH in a query. For this, like in [Cima et al. \(2017\)](#), we have introduced in both LUBM(1) and (9) a new class named *TypeOfProfessor* and make *FullProfessor*, *AssociateProfessor* and *AssistantProfessor* an instance of this new class, which makes *TypeOfProfessor* a meta-

³ <http://swat.cse.lehigh.edu/projects/lubm/queries-sparql.txt>.

⁴ The reason for this naming convention is, these queries are specifically created to check the meta-modeling case in LUBM-ext ontology.

class and also we define *FullProfessor*, *AssociateProfessor* and *AssistantProfessor* to be disjoint from each other. Then, in *sq1* we asked for all those y and z , where y is a professor, z is a type of professor and y is an instance of z , violating the typing constraint. In *sq2*, we asked for instances of type of professor that are disjoint to see the impact of negative term for the evaluation of the query. We referred the extension of LUBM ontologies with additional expressivity as LUBM-ext, LUBM(1)-ext and LUBM(9)-ext.

The MODEUS dataset comes with 9 queries.⁵ These are all meta-queries, containing variables referring to more than one type, these queries traverse over several layers of *instanceOf* relation. These queries represent the chain of *instanceOf* relation, where some chains are expressed as *classAssertion*, start with a specific entity, start with *subclass*, asking for meta-property and some constituted by joining more than one chains of *instanceOf*.

The queries *mq1*, *mq4*, *mq5*, and *mq10* are simpler forms of meta-queries that do not span over different levels of ontologies, as the variables appearing as class and property names in these queries do not appear in other types of positions. The queries *sq1*, *sq2* and the 9 MODEUS queries are pure forms of the meta-queries that span over multiple level of ontologies.

4.7.1 Test environment

We have done the test on a Linux batch server, running Ubuntu 20.04.3 LTS (GNU/Linux 5.4.0-88-generic x86_64) on one AMD EPYC 7601 (32-Core CPU), 2.2GHz, Turbo max. 3.2GHz. The machine is equipped with 515GB RAM and a 4TB hard disk. Java applications used OpenJDK 11.0.11 with a maximum heap size of 25GB.

To perform different settings and measurements on the system resources we have used *Pyrunlim*⁶ and *BenchExec*.⁷ The reason for using two different tools for resource limitation is due to *LogicBlox*, because it creates spawned processes, which causes resource consumption to be partially unaccounted for with *Pyrunlim*. So, to limit the spawned processes on Linux, we have used the *cgroups* features of resource limitation and measurement provided by *BenchExec*. Since we ran the *LogicBlox* experiments last and encountered this issue, we did not re-run the other experiments with *BenchExec*, as there should not be any dependence on the measurement tool. We manually verified on selected examples that *Pyrunlim* and *BenchExec* yield the same results up to measurement precision.

4.7.2 Performance breakdown

During the course of the evaluation of Datalog-based MSER we have used the two different resource limitations (on RAM and Time) as the benchmark setting on our data sets to examine the behavior of different tools with tight or ample resources. Thus, we can provide additional insight on the specific gains obtained by integrating various settings. We report the time of computing the answers to each query and include loading

⁵ <http://www.modeus.uniroma1.it/modeus/node/6>.

⁶ <https://github.com/alviano/python/tree/master/pyrunlim>.

⁷ <https://github.com/sosy-lab/benchexec>.

time. For simplicity, we have not included queries that contain the data properties in our experiments.

If not otherwise indicated, in the first experiment, each benchmark had 15 min and 8GB of memory on a Linux cluster. We report the time of computing the answer sets for each query. In the second experiment, we increased the time to 30 min and 128GB of memory.

5 Experimental results

We now report the experimental results under the two resource bounds, starting with the tighter limits, followed by the more generous ones.

5.1 Experiment 1 – tight resource limits

We next report the results on different datasets under the first experiment setting, first LUBM with standard queries,⁸ followed by meta-queries over LUBM, then moving towards MODEUS with meta-queries.

In the first batch of experiments for each tool, we have limited the RAM usage to 8GB and allowed for a maximum of 15 min to answer the query. In Figure 1 we have reported the results for different ABox sizes of the LUBM ontologies. All times reported in these tables are in seconds and include loading the Datalog program including facts and rules and answering the query. The best performance for each query is highlighted in boldface. The value *OOT* and *OOM* refers to *Out-Of-Time* and *Out-Of-Memory*.

5.1.1 LUBM with standard-queries and simple meta-queries

We can observe that performance is satisfactory for LUBM(1). Of these 14 queries, q2, q7, and q9 are the most interesting ones since they contain many atoms affecting running time. In these queries, we observe an increase in running time for some tools. We can see, however, that LogicBlox, NoHR, HexLite, and Alpha generally introduce significant overhead, which we did not really expect on this scale.

It is not fully clear to us yet what causes this; the ontology part is empty in the hybrid system NoHR, so we would have expected NoHR performance to be similar to XSB. It might be that the tabling strategy is to blame in LUBM(1) and in LUBM(9), we have observed that time and tabling strategy is directly proportional to each other, but this is just a conjecture. There is one instance in which XSB is quite slow (q9 on LUBM(1) and LUBM(9)), but performance on this query is generally bad and a different indexing strategy might be the cause for this behaviour. For HexLite, we would have expected more or less the same performance to DLVHex/Clingo as the queries were simple, there is not much external computation involved, and it uses Clingo for searching and grounding. It might be possible that the external OWL-API plugin that we used with HEXLite caused an overhead for HEXLite. As for Alpha, it is not an optimised

⁸ As one of the reviewer pointed out – this setting is interesting that it allows to compare the cost of evaluating meta-queries w.r.t standard queries using different Datalog engines, over an ontology that does not contain any meta-predicate.

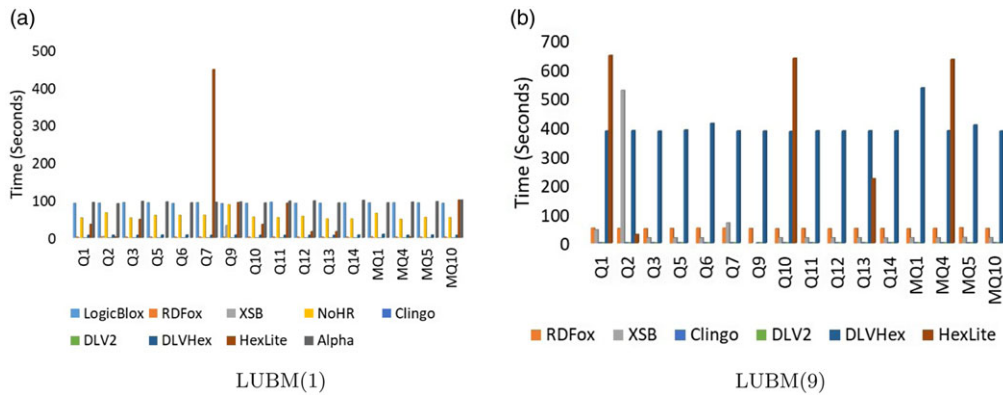


Fig. 1. Experiment 1 LUBM with standard queries and simple meta-queries.

system, but we were expecting better performance than other ASP tools. Furthermore, Alpha performance degraded in LUBM(9) when the input data size increases and we could not observe benefits of lazy grounding in our experiments.

Across all queries of LUBM(1), LogicBlox exhibits very stable performance with roughly the same execution time for all queries, which is quite remarkable. But it lost this performance on LUBM(9), where the ontology size has affected its performance. Also, DLV2 and Clingo have very regular performance, but the time is affected slightly by the size of the dataset. RDFox is also very regular across different queries but is quite a bit affected by the size of the dataset. DLVHex, HexLite, and XSB show varying performance with some queries over the same ontology. While XSB is the fastest and consistent system on some queries over LUBM(1), it is among the slowest on other queries. This becomes more pronounced for LUBM(9), where XSB is still quick for a few queries but really slow on others. DLVHex is quicker on LUBM(1) than HEXLite, which shows the contradictory performance on both ontologies.

5.1.2 LUBM-ext with meta-queries

It can be seen in Figure 2 that the overall performance of meta-query evaluation is similar to the one in Figure 1. LogicBlox has again the most regular performance in LUBM(1)-ext but shows performance deterioration for LUBM(9)-ext. Clingo and DLV2 have regular performance over varying queries over the same ontology, but are slightly affected by the ontology size. Clingo and DLV2 have shown best performance on a query sq1, next to XSB. Similar comments apply to RDFox and XSB (especially on LUBM(9)-ext), but they are overall slower and performance deteriorates more. XSB shows the best and consistent performance on LUBM(1). NoHR shows somewhat more regular performance here to the one in Figure 1, but it deteriorates with the ontology size and suffers because of the time restriction.

DLVHex shows the same performance on LUBM(1)-ext similar to the results in Figure 1, but its performance gets more affected in LUBM(9)-ext. In contrast, HEXLite, performance suffers from time limits. Alpha shows more or less the same performance on LUBM(1)-ext than in Figure 1 but was not able to stay within the memory limit on LUBM(9)-ext.

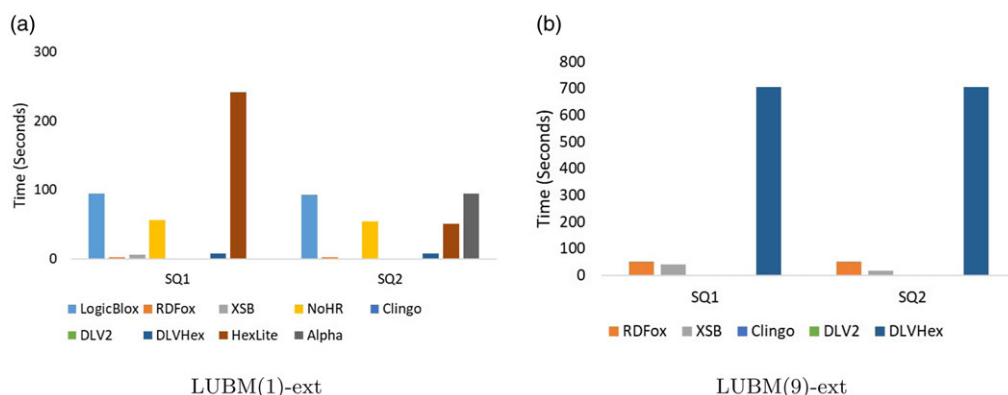


Fig. 2. Experiment 1 LUBM-ext with meta queries.

5.1.3 MODEUS with meta-queries

The results are reported in Figure 3. Except for mq4 and mq8, for the rest of the queries mq0-mq6, we observe the number of individuals in the ontologies, and intermediate results (determine the next query atom to be executed in each iteration) are quite large. The cause of iterating over all the possible mappings constituted by an *InstanceOf* chain > 1 causes an overhead of memory and running time. We can see that most of the tools were not able to compute answers with the given resources. Some tools ran out of memory while loading the data, like LogicBlox and RDFox. In contrast, others consume a lot of memory while grounding atoms like Clingo, DLVHex, HEXLite, and Alpha. NoHR exhibits varied performance, where it successfully computed some queries in the given time, while it ran out of time for others, apparently due to tabling atoms. However, NoHR suffers from quite severe overhead with MODEUS queries compared to the LUBM queries. From the results of Figure 3, one can observe that the queries mq4 and mq8 perform better than others because both queries start with the specific entity and are constituted by an *InstanceOf* chain of length 1.

Other than the resource limitations, there are also some considerations concerning the nature of the MODEUS ontologies:

- As mentioned earlier, the MODEUS dataset consists of meta-layers, which appear to cause many tools to do more inferencing;
- the meta-queries consist of different layers of classes, instances and properties, which span over several layers of the dataset; and
- we conjecture that the presence of many *disjoint* axioms causes particularly many inferences.

XSB and DLV2 show regular performance throughout each dataset. DLV2 is the best overall performer in Figure 3. Its performance is very regular and consistently quick. We assume that the magic set technique implemented in DLV2 has a huge impact here, but it is not clear why LogicBlox (which is also supposed to implement one) cannot profit from it. A possibility is that different sideways information passing strategies (SIPS) are used, which might explain this behavior.

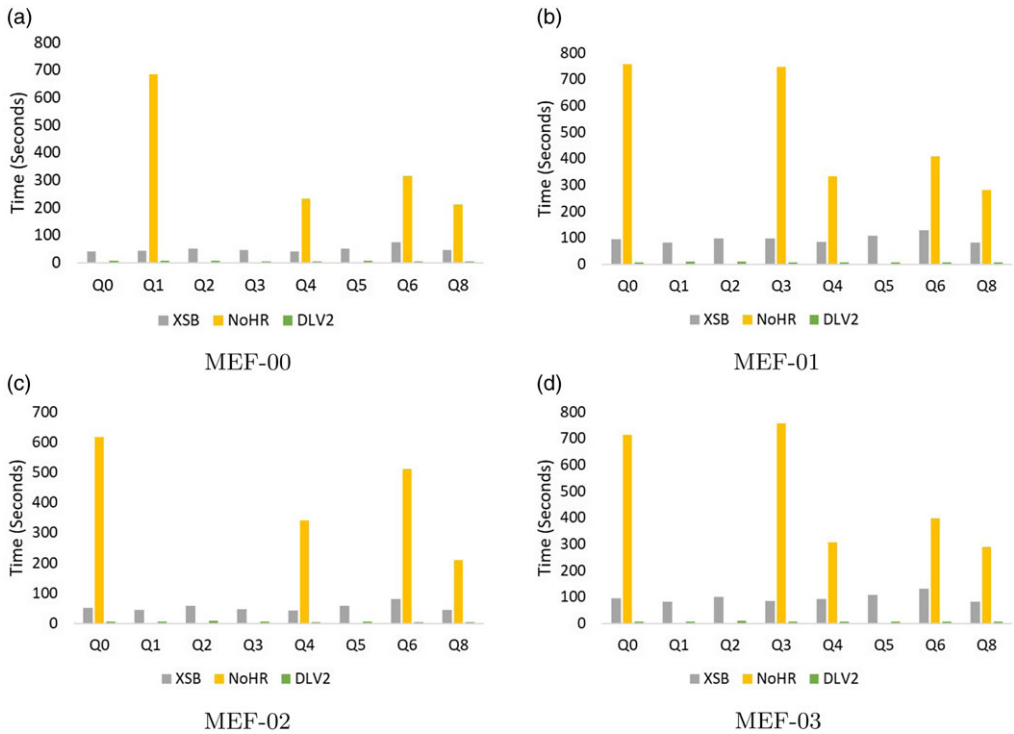


Fig. 3. Experiment 1 MODEUS with meta queries.

XSB also shows comparatively regular performance but less so than DLV2. It is interesting to see that XSB cannot profit more from its top-down strategy; it is possible that tabling is actually slowing it down in these examples.

5.2 Experiment 2 – more generous resource limits

We next report the results on different datasets under the second experiment settings, allowing for 30 min and 128GB first LUBM with standard queries following LUBM-ext with meta-queries, then moving toward MODEUS with meta-queries.

5.2.1 LUBM with standard queries and simple meta-queries

It can be seen in Figure 4(a) and (b) that almost all tools were able to compute answers with the given increased resources, except for XSB with one outlier (q9) that ran out-of-time during tabling, HexLite with five outliers (q5, q6, q14, mq1, and mq5) in LUBM(1), nine (q5, q6, q7, q9, q11, q14, mq1, mq5, and mq10) in LUBM(9), all of which were out-of-time during the search phase, and NoHR ran out-of-time during preprocessing of all queries in LUBM(9).

DLVHex exhibits the same performance as in Experiment 1 in Figure 1, whereas some queries in HexLite still exceeded the time limits. Alpha has shown the same performance as in Experiment 1 in Figure 1 on LUBM(1) and is able to meet the memory limit on LUBM(9), but the performance was not particularly satisfactory.

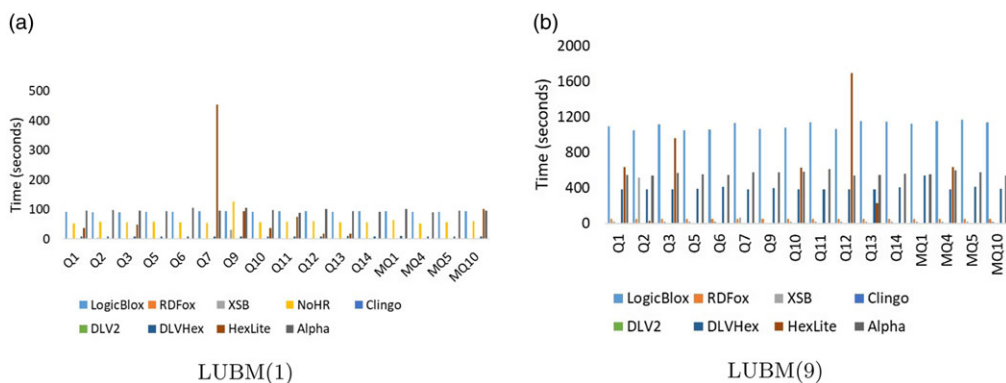


Fig. 4. Experiment 2 LUBM with standard queries and simple meta-queries.

Experiment 2 confirmed the observations for most tools of experiment 1. In particular, XSB, DLV2, and Clingo show consistent performance, where XSB outperforms every single tool in LUBM(1) but is affected by the ontology size in LUBM(9); on the other hand, Clingo and DLV2 show regular performance across both ontologies with a slight increase in time for LUBM(9).

LogicBlox benefits significantly from increased resources. It now shows consistent performance for both LUBM(1) and LUBM(9). However, it can be seen that its performance deteriorates with the size of the ontology. RDFox shows good performance in LUBM(1) but not so much for LUBM(9). Nevertheless, the overall performance of RDFox across both LUBM(1) and LUBM(9) in both experiments is consistent, which is quite remarkable.

5.2.2 LUBM-ext with meta-queries

It can be seen in Figure 5(a) and (b) that HexLite ran out-of-time during the execution of sq1 and NoHR ran out-of-time during the execution of sq1 and sq2 in LUBM(9)-ext. Both tools, HexLite and NoHR, performance greatly affected by the size of the ontologies. Other than the previous two, Alpha performance get very much affected by the ontology size. However, it is able to evaluate both queries but not the quickest.

LogicBlox, performance deteriorates with the extension and we have observed that the "disjoint-axiom" has a significant impact on the performance as it seems to cause LogicBlox to check and apply many alternatives. RDFox shows good performance in LUBM(1)-ext, but its performance is affected by the ontology size in LUBM(9)-ext. But overall the RDFox performance is satisfactory on the LUBM-ext.

XSB, Clingo and DLV2 are quickest on LUBM(1)-ext and exhibit regular performance. However, XSB performance is slightly affected by the size of the input ontology but still performs better than RDFox. Clingo and DLV2 show the most regular and consistent performance over both ontologies and DLV2 outperforms all the tools.

5.2.3 MODEUS with meta-queries

In Experiment 2, we get some interesting results compared to Experiment 1. First, XSB, DLV2, and HexLite show a similar performance. Second, the tools that broke the memory

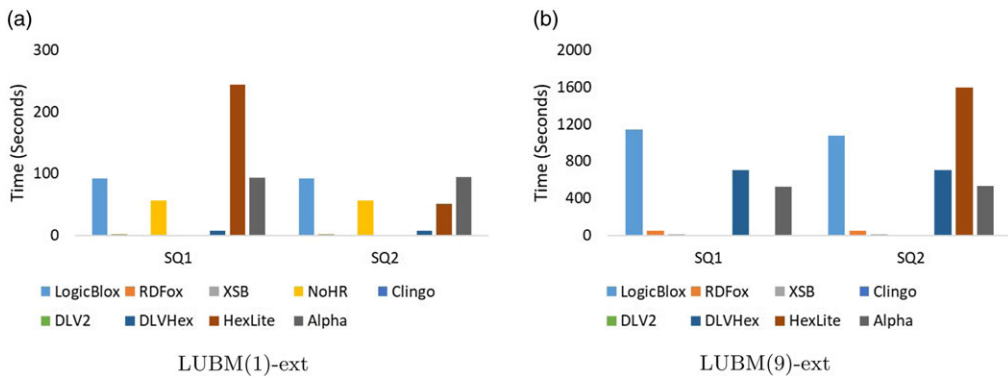


Fig. 5. Experiment 2 LUBM-ext with meta queries.

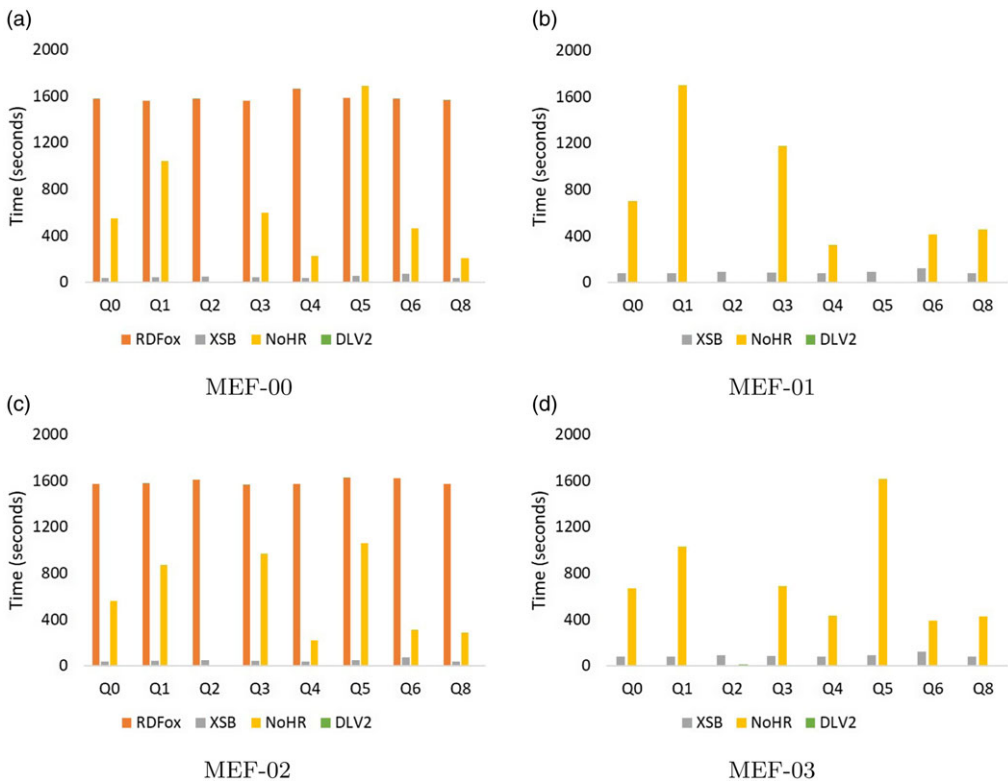


Fig. 6. Experiment 2 MODEUS with meta queries.

limits in Experiment 1 like LogicBlox, DLVHex, and Alpha also suffered from the time limit in this experiment. Third, NoHR that could only answer some queries in Experiment 1 was able to answer almost all queries over all four MODUES ontologies except for q2 in MEF-00, MEF-02, MEF-03, and q2,q5 in MEF-01.

Fourth, RDFox shows its performance on MEF-00 and MEF-02 and suffers from time limitations on the remaining two ontologies. However, the performance on the former two ontologies was not the best performance. Lastly, Clingo produces an error while

grounding, indicating that the system ran out of identifiers for ground atoms. Apparently, the problem gives rise to a ground program that is too large for Clingo to handle.

DLV2 and XSB show the most consistent performance across all experiments. However, DLV2 outperforms every single tool in both experiments. The magic set technique appears to play a significant role in the performance of DLV2. On the other hand, the tabling and top-down querying approach have played their role in the consistent performance of XSB across different ontologies and experiments.

6 Discussion and future work

In this paper, we provided both a brief background and an evaluation of a method for answering SPARQL queries over OWL 2 QL ontologies under the MS via a reduction to Datalog queries. The main contribution of our paper is a performance analysis of various tools that support Datalog, significantly expanding the preliminary analysis in [Cima *et al.* \(2017\)](#): on the one hand, we additionally consider the MODEUS ontologies that involve meta-axioms, and on the other hand, we evaluate significantly more tools of different backgrounds that support Datalog query answering.

Indeed, our experiments show that especially DLV2, but also XSB appear to be promising back-ends for meta-querying over OWL 2 QL ontologies containing meta-modeling. Surprisingly, neither RDFox nor LogicBlox coped very well with the MODEUS ontologies that involve meta-modeling, while LogicBlox performed very well on LUBM(1) queries. Also, RDFox does not seem to be a good fit for this particular kind of application. Also, the overhead created by NoHR appears quite prohibitive in this scenario. Lastly, Clingo, Alpha, DLVHex, and HexLite do not seem to be well-suited for query answering of this particular kind.

Next, we provide our observations on some tools and try to explain why there is a difference between their performance.

Contrasted with the best in class grounding and solving frameworks like Clingo, Alpha is not yet on par when the grounding is relatively simple to construct, and search is comparatively complex. In our benchmark, we would have expected Alpha to perform well, as there is little search to be done. However, it seems that Alpha is not really geared towards query answering scenarios. We believe that there is potential for Alpha to become a dedicated query answering engine as well.

On the other hand, the limits of Clingo's architecture became apparent with the MODEUS ontologies. Clingo integrates the grounder gringo and the solver clasp into a monolithic system to offer more control over the grounding and solving process. Clasp has an inbuilt limit of 2^{28} atoms, triggering an error with the MODEUS ontologies, while gringo can handle up to 2^{30} atoms. It is apparently not possible to simply increase these limits. Clingo could benefit from a magic set technique as employed by DLV2, which would avoid these size issues.

Compared to the state-of-the art HEX solver system DLVHex, HexLite is not yet competitive on problems of external atoms computations. HexLite appears to pass over all the optimization tasks (to check the consistency of assignments) to the backend solver Clingo and, as discussed above, Clingo has an inbuilt limit for ground atoms.

DLVHex, which implements the complete Hex language shows quite poor performance on the MODEUS ontologies; it performs a lot of analysis (creating guesses for all external

atom even if they are not relevant), bookkeeping and additional computation for optimization to deal with all occurrences of combinations of external computations, which makes DLVHex unnecessarily slow in MODEUS ontologies.

Related to this work are Poggi (2016), Kontchakov *et al.* (2014), Gottlob and Pieris (2015), which consider OWL 2 QL ontologies under DSER and show that it is decidable and can be reduced to Datalog evaluation. However, those works do not consider meta-modeling and meta-querying.

In Lenzerini *et al.* (2016a, 2015), the authors overcome the limitations of querying over OWL 2 QL ontologies by introducing the MS and show that their proposed algorithm tackles the untyped querying problem in PTime w.r.t. data and ontology complexity. The proposed algorithm is implemented in MQ-Mastro,⁹ an OWL 2 QL reasoner. It allows for meta-querying over meta-modeling ontologies by offering two types of query evaluation algorithms, NAIVE and Lazy Meta-Grounding (LMG) Lenzerini *et al.* (2020). The system comes with a Graphical User Interface (GUI). Unfortunately, we did not succeed to run it in batch mode, but loading and preprocessing took several minutes, while answering queries was quick. Overall, however, it cannot match performance of the best Datalog systems.

As future work we intend to further examine hybrid reasoning for answering meta-queries. The main idea would be to use the same translation function τ and inference rules, but to keep the non-meta-modeling parts of the ontology untranslated. This would allow for the use of standard ontology reasoners for the latter part and could lead to performance benefits. The work presented in this paper would be an extreme case of this setting, in which the untranslated part is always empty.

Supplementary material

To view supplementary material for this article, please visit <http://doi.org/10.1017/S1471068424000073>.

References

- ALVIANO, M., CALIMERI, F., DODARO, C., FUSCÀ, D., LEONE, N., PERRI, S., RICCA, F., VELTRI, P. AND ZANGARI, J. 2017. The ASP system DLV2. In *Logic Programming and Non-monotonic Reasoning: 14th International Conference, LPNMR 2017, Espoo, Finland, July 3–6, 2017, Proceedings 14*. Springer, 215–221.
- ANTONIOU, G. AND VAN HARMELEN, F. 2009. Web ontology language: OWL. In *Handbook on Ontologies, International Handbooks on Information Systems*. Springer, 91–110.
- AREF, M., TEN CATE, B., GREEN, T. J., KIMELFELD, B., OLTEANU, D., PASALIC, E., VELD-HUIZEN, T. L. AND WASHBURN, G. 2015. Design and implementation of the LogicBlox system. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1371–1382.
- BAADER, F., CALVANESE, D., MCGUINNESS, D., PATEL-SCHNEIDER, P., NARDI, D. 2003. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press.

⁹ <http://www.modeus.uniroma1.it/modeus/node/6>.

- BRASILEIRO, F., ALMEIDA, J. P. A., CARVALHO, V. A. AND GUIZZARDI, G. 2016. Expressive multi-level modeling for the semantic web. In *International Semantic Web Conference*. Springer, 53–69.
- BREWKA, G., EITER, T. AND TRUSZCZYŃSKI, M. 2011. Answer set programming at a glance. *Communications of the ACM* 54, 12, 92–103.
- BRUIJN, J. D., EITER, T., POLLERES, A. AND TOMPITS, H. 2011. Embedding nonground logic programs into autoepistemic logic for knowledge-base combination. *ACM Transactions on Computational Logic (TOCL)* 12, 3, 1–39.
- CALVANESE, D., DE GIACOMO, G., LEMBO, D., LENZERINI, M. AND ROSATI, R. 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning* 39, 3, 385–429.
- CERI, S., GOTTLOB, G. AND TANCA, L. 2012. *Logic Programming and Databases*. Springer Science & Business Media.
- CHEN, W., KIFER, M. AND WARREN, D. S. 1993. HiLog: A foundation for higher-order logic programming. *The Journal of Logic Programming* 15, 3, 187–230.
- CIMA, G., DE GIACOMO, G., LENZERINI, M. AND POGGI, A. 2017. On the SPARQL meta-modeling semantics entailment regime for OWL 2 QL ontologies. In *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics*, 1–6.
- DAO-TRAN, M., EITER, T., FINK, M., WEIDINGER, G. AND WEINZIERL, A. 2012. Omega: An open minded grounding on-the-fly answer set solver. In *European Workshop on Logics in Artificial Intelligence*. Springer, 480–483.
- DE MOOR, O., GOTTLOB, G., FURCHE, T. AND SELLERS, A. 2012. *Datalog Reloaded: First International Workshop, Datalog 2010, Oxford, UK, March 16–19, 2010. Revised Selected Papers*, Vol. 6702. Springer.
- DE MOOR, O., SERENI, D., AVGUSTINOV, P. AND VERBAERE, M. 2008. Type inference for datalog and its application to query optimisation. In *Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 291–300.
- EITER, T., IANNI, G., LUKASIEWICZ, T., SCHINDLAUER, R. AND TOMPITS, H. 2008. Combining answer set programming with description logics for the semantic web. *Artificial Intelligence* 172, 12–13, 1495–1539.
- EITER, T., IANNI, G., SCHINDLAUER, R. AND TOMPITS, H. 2006. dlhex: A prover for semantic-web reasoning under the answer-set semantics. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*. IEEE, 1073–1074.
- EITER, T. AND KRENNWALLNER, T. 2012. *Reasoning Web-Semantic Technologies for Advanced Query Answering: 8th International Summer School 2012, Vienna, Austria, September 3–8, 2012. Proceedings*, Vol. 7487. Springer.
- GEBSER, M., KAMINSKI, R., KAUFMANN, B., OSTROWSKI, M., SCHAUB, T. AND THIELE, S. 2010. gringo, clasp, clingo, and iclingo.
- GEBSER, M., KAUFMANN, B. AND SCHAUB, T. 2012. Conflict-driven answer set solving: From theory to practice. *Artificial Intelligence* 187, 52–89.
- GLIMM, B. 2011. Using SPARQL with RDFs and OWL entailment. In *Reasoning Web International Summer School*. Springer, 137–201.
- GOTTLOB, G. AND PIERIS, A. 2015. Beyond SPARQL under OWL 2 QL entailment regime: Rules to the rescue. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- GUIZZARDI, G., ALMEIDA, J. P. A., GUARINO, N. AND DE CARVALHO, V. A. 2015. Towards an ontological analysis of powertypes. In *JOWO@IJCAI*.
- GUO, Y., PAN, Z. AND HEFLIN, J. 2005. LUBM: A benchmark for OWL knowledge base systems. *Journal of Web Semantics* 3, 2–3, 158–182.

- HORROCKS, I., PARSIA, B. AND SATTLER, U. 2012. OWL 2 web ontology language direct semantics. In *World Wide Web Consortium*, 42–65.
- HORROCKS, I., PATEL-SCHNEIDER, P. F. AND VAN HARMELEN, F. 2003. From SHIQ and RDF to OWL: The making of a web ontology language, *Journal of Web Semantics* 1, 1, 7–26.
- HUANG, S. S., GREEN, T. J. AND LOO, B. T. 2011. Datalog and emerging applications: An interactive tutorial. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, 1213–1216.
- KONTCHAKOV, R., REZK, M., RODRIGUEZ-MURO, M., XIAO, G. AND ZAKHARYASCHEV, M. 2014. Answering SPARQL queries over databases under OWL 2 QL entailment regime. In *International Semantic Web Conference*. Springer, 552–567.
- LENZERINI, M., LEPORE, L. AND POGGI, A. 2015. A higher-order semantics for OWL 2 QL ontologies. *Description Logics*.
- LENZERINI, M., LEPORE, L. AND POGGI, A. 2016a. Answering metaqueries over Hi (OWL 2 QL) ontologies. In *IJCAI*, 1174–1180.
- LENZERINI, M., LEPORE, L. AND POGGI, A. 2016b. A higher-order semantics for metaquerying in OWL 2 QL. In *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*.
- LENZERINI, M., LEPORE, L. AND POGGI, A. 2020. Metaquerying made practical for OWL 2 QL ontologies. *Information Systems* 88, 101294.
- LENZERINI, M., LEPORE, L. AND POGGI, A. 2021. Metamodeling and metaquerying in OWL 2 QL. *Artificial Intelligence* 292, 103432.
- LLOYD, J. W. 2012. *Foundations of Logic Programming*. Springer Science & Business Media.
- LOPES, C., KNORR, M. AND LEITE, J. 2017. Nohr: Integrating XSB prolog with the OWL 2 profiles and beyond. In *Logic Programming and Nonmonotonic Reasoning: 14th International Conference, LPNMR 2017, Espoo, Finland, July 3–6, 2017, Proceedings 14*. Springer, 236–249.
- MOTIK, B. 2007. On the properties of metamodeling in OWL. *Journal of Logic and Computation* 17, 4, 617–637.
- MOTIK, B., GRAU, B. C., HORROCKS, I., WU, Z., FOKOUE, A., LUTZ, C. 2009. OWL 2 web ontology language profiles. *W3C Recommendation* 27, 61.
- MOTIK, B. AND ROSATI, R. 2007. A faithful integration of description logics with logic programming. In *IJCAI*, Vol. 7, 477–482.
- NENOV, Y., PIRO, R., MOTIK, B., HORROCKS, I., WU, Z. AND BANERJEE, J. 2015. RDFox: A highly-scalable RDF store. In *The Semantic Web-ISWC 2015: 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11–15, 2015, Proceedings, Part II 14*. Springer, 3–20.
- PÉREZ, J., ARENAS, M. AND GUTIERREZ, C. 2009. Semantics and complexity of SPARQL. *ACM Transactions on Database Systems (TODS)* 34, 3, 1–45.
- POGGI, A. 2016. On the SPARQL direct semantics entailment regime for OWL 2 QL. In *Description Logics*.
- QURESHI, H. M. AND FABER, W. 2021. An evaluation of meta-reasoning over OWL 2 QL. In *International Joint Conference on Rules and Reasoning*. Springer, 218–233.
- SCHÜLLER, P. 2019. The hexlite solver: Lightweight and efficient evaluation of HEX programs. In *European Conference on Logics in Artificial Intelligence*. Springer, 593–607.
- SWIFT, T. AND WARREN, D. S. 2012. XSB: Extending prolog with tabled logic programming. *Theory and Practice of Logic Programming* 12, 1-2, 157–187.
- WANG, K., BILLINGTON, D., BLEE, J. AND ANTONIOU, G. 2004. Combining description logic and defeasible logic for the Semantic Web. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*. Springer, 170–181.

- WEINZIERL, A. 2017. Blending lazy-grounding and CDNL search for answer-set solving. In *International Conference on Logic Programming and Nonmonotonic Reasoning*. Springer, 191–204.
- WEINZIERL, A., BOGAERTS, B., BOMANSON, J., EITER, T., FRIEDRICH, G., JANHUNEN, T., KAMINSKI, T., LANGOWSKI, M., LEUTGEB, L., SCHENNER, G. AND TAUPE, R. 2019. The Alpha solver for lazy-grounding answer-set programming. *ALP Newsletter*.