

The (Not So) Easy Task of Computing Class Subsumptions in OWL RL^{*}

Technical Report

Markus Krötzsch

Department of Computer Science, University of Oxford, UK,
`markus.kroetzsch@cs.ox.ac.uk`

Abstract. The lightweight ontology language OWL RL is used for reasoning with large amounts of data. To this end, the W3C standard provides a simple system of deduction rules, which operate directly on the RDF syntax of OWL. Several similar systems have been studied. However, these approaches are usually complete for instance retrieval only. This paper asks if and how such methods could also be used for computing entailed subclass relationships. Checking entailment for arbitrary OWL RL class subsumptions is co-NP-hard, but tractable rule-based reasoning is possible when restricting to subsumptions between atomic classes. Surprisingly, however, this cannot be achieved in any RDF-based rule system, i.e., the W3C calculus cannot be extended to compute all atomic class subsumptions. We identify syntactic restrictions to mitigate this problem, and propose a rule system that is sound and complete for many OWL RL ontologies.

1 Introduction

The lightweight ontology language OWL RL [16] is widely used for reasoning with large amounts of data, and many systems support query answering over OWL RL ontologies. Commercial implementations of (parts of) OWL RL include Oracle 11g [13], OWLIM [2], Virtuoso [5], and AllegroGraph [6].

What makes OWL RL so appealing to users and implementers alike are its favourable computational properties. As for all three lightweight profiles of OWL 2, typical reasoning tasks for OWL RL can be solved in polynomial time. Possibly even more important, however, is the fact that this low complexity can be achieved by relatively straightforward algorithms that perform bottom-up, rule-based materialisation of logical consequences until saturation. While it is also possible to use similar algorithms for OWL EL [15], a simple rule-based algorithm for OWL RL is already given in the W3C specification [16]. Various similar rule sets have been published for fragments of OWL RL (e.g., [10,8,21]).

Another advantage of these rule systems is that they operate directly on the RDF serialisation of OWL. OWL RL reasoning thus can easily be implemented

^{*} Revised version of 13 Nov 2012. See end of document for revision history.

on top of existing RDF databases that support some form of production rules to infer additional information. Basic rule-matching capabilities are found in most RDF stores, since they are similar to query matching. Indeed, *SPARQL rules* have been proposed as a natural rule extension of the SPARQL query language that can fully express the OWL RL rule system [20]. It is important that neither value invention (*blank node creation*) nor non-monotonic negation are needed in OWL RL, as both features complicate rule evaluation significantly [20].

A common strategy for evaluating larger amounts of OWL RL data is to separate terminological information (schema-level axioms) from assertional information (facts), since the latter are typically significantly larger [14,9,21]. To further reduce the rules that need to be applied to the data, it would be useful to pre-compute terminological inferences, especially all subclass relationships.

Unfortunately, it is not known how to do this. To the best of our knowledge, no polynomial time algorithm has been published for computing schema entailments in OWL RL. As shown in Section 2, the W3C rule system is not complete for computing class subsumptions, and, more problematically, it is not possible to compute all class subsumptions without taking assertions into account.

However, it is still possible to obtain RDF-based rule systems that can discover more entailments than the W3C rules. Using various abbreviations introduced in Section 3, we present one such algorithm in Section 4. We identify `ObjectHasValue` as the main obstacle – if it does not occur in superclasses, our algorithm can compute all class subsumptions. In Section 5, we take a look at RDF-based rules and show how our results transfer to this setting.

In Section 6, we discuss the problem of computing class subsumptions in unrestricted OWL RL ontologies, where `ObjectHasValue` is allowed in superclasses. It is still possible to use polynomial time rule systems for this case, but it turns out that there is no *RDF-based* rule system for this task. This surprising result shows an inherent limitation of the expressive power of RDF-based rules.

Most of our presentation is based on the Functional-Style Syntax of OWL [17]. This yields a more concise presentation (compared to the RDF serialisation) and still is close to the actual language. We assume basic familiarity with the syntax and semantics of OWL on the level of the *OWL Primer* [7]. If not stated otherwise, we generally consider the Direct Semantics of OWL, but we also mention some results about the RDF-based Semantics. When writing OWL entities in axioms, we tacitly use prefixes for suggesting abbreviated IRIs.

2 Challenges of Schema Reasoning in OWL RL

Before we focus on the task of computing class subsumptions for OWL RL, it is worth pointing out some limitations and challenges that do not seem to be well known, even among practitioners and implementers.

Checking Entailment of OWL RL Axioms is Not Tractable The W3C specification mentions that *Class Expression Subsumption* in OWL RL is PTime-complete w.r.t. the size of the ontology [16, Section 5]. This might evoke the

impression that one could check in polynomial time whether an OWL RL class inclusion axiom is entailed by an OWL RL ontology. This is not the case.¹

Proposition 1. *Given an OWL RL ontology \mathcal{O} and an OWL RL `SubClassOf` axiom A , checking whether \mathcal{O} entails A is co-NP-hard.*

Proof. We show this by reducing 3SAT unsatisfiability to OWL RL entailment checking. An instance of the 3SAT problem is a set of propositional clauses $\{(L_{11} \vee L_{12} \vee L_{13}), \dots, (L_{n1} \vee L_{n2} \vee L_{n3})\}$, where each L_{ij} is a propositional variable or a negated propositional variable. The question whether the conjunction of these clauses is satisfiable is NP-complete. For each propositional variable p , we introduce two new class names T_p and F_p . To each literal L_{ij} , we assign a class name $c(L_{ij})$ as follows: if $L_{ij} = p$, then $c(L_{ij}) := T_p$; if $L_{ij} = \neg p$, then $c(L_{ij}) := F_p$. For every clause $(L_{i1} \vee L_{i2} \vee L_{i3})$, we define a class expression C_i as `ObjectUnionOf(c(L11) c(L12) c(L13))`. Now let A be the axiom `SubClassOf(ObjectIntersectionOf(C1 ... Cn) owl:Nothing)`, and let \mathcal{O} be the ontology that consists of the axioms `DisjointClasses(Tp Fp)` for every propositional variable p . Clearly, A is an OWL RL axiom and \mathcal{O} is an OWL RL ontology. However, \mathcal{O} entails A if and only if the given instance of 3SAT has *no* solution. Indeed, if A is not entailed, then \mathcal{O} has a model where an individual e is an instance of each of the class expression C_i . We can construct a propositional truth assignment as follows: if e is an instance of T_p , then p is mapped to true; otherwise p is mapped to false. It is easy to see that this is a solution to the 3SAT instance, since e cannot be an instance of T_p and F_p for any p . \square

Another way to find hardness proofs is to use `DataSomeValuesFrom` in the subclass, together with datatypes such as XML Schema `boolean`, which is admissible in OWL RL subclasses. Moreover, similar problems can be found for other axiom types; e.g., checking entailment of `hasKey` axioms is also intractable.

These problems are hardly surprising from a logical perspective, since checking the entailment of A from \mathcal{O} is equivalent to checking the consistency of $\mathcal{O} \cup \{\neg A\}$, where $\neg A$ is the negation of the axiom A (as a logical formula). For the latter to be in OWL RL, one needs to impose different syntactic restrictions on A . The check is then possible in PTime. A particularly relevant case where this is possible is that A is a subclass relationship between two class names. This is the task that we will focus on in the remainder of this work.

The W3C Rule System is Not Complete for Class Subsumption Checking The W3C specification states a completeness theorem for its rule system, which asserts completeness only for entailment of assertional axioms. However, the rule system contains a number of rules that would not be necessary to achieve this especially in Table 9, entitled *The Semantics of Schema Vocabulary*. This might lead to the wrong impression that the rules can infer all schema axioms, or at least all class subsumptions. The following example illustrates that this is wrong.

¹ The 2012 update of the OWL 2 specification will correct this; see Section 7.

Example 1. We consider an ontology of three axioms:

$$\text{SubClassOf}(:A :B) \tag{1}$$

$$\text{SubClassOf}(:A :C) \tag{2}$$

$$\text{SubClassOf}(\text{ObjectIntersectionOf}(:B :C) :D) \tag{3}$$

This ontology clearly entails $\text{SubClassOf}(:A :D)$. However, this is not entailed by the W3C rule system. The only entailment rules that refer to $\text{ObjectIntersectionOf}$ are rules *cls-int1*, *cls-int2*, and *scm-int* in [16]. The former two rules are only applicable to individual instances. Rule *scm-int* can be used to infer $\text{SubClassOf}(\text{ObjectIntersectionOf}(:B :C) :B)$ and $\text{SubClassOf}(\text{ObjectIntersectionOf}(:B :C) :C)$ – the rule can be viewed as a schema-level version of *cls-int2*. However, there is no rule that corresponds to *cls-int1* on the schema level, so one can not infer $\text{SubClassOf}(:A \text{ObjectIntersectionOf}(:B :C))$.

This example extends to many other types of axioms. For example, one cannot infer all entailed property domains or ranges if some class subsumptions have been missed.

Assertions Can Not be Ignored when Checking Class Subsumptions Since many OWL RL ontologies are dominated by assertional axioms (i.e., data), it would be useful if this part of the ontology would not be relevant if one is only interested in computing class subsumptions. This cannot work in general, since assertions can cause the ontology to become inconsistent, which in turn leads to the entailment of arbitrary class subsumptions. However, even for consistent ontologies, assertions cannot be ignored when computing class subsumptions, as shown in the next example.

Example 2. We consider an ontology of three axioms:

$$\text{ClassAssertion}(:B :b) \tag{4}$$

$$\text{SubClassOf}(:A \text{ObjectHasValue}(:P :b)) \tag{5}$$

$$\text{SubClassOf}(\text{ObjectSomeValuesFrom}(:P :B) :C) \tag{6}$$

This ontology entails $\text{SubClassOf}(:A :D)$: every instance of $:A$ has a $:P$ successor $:b$ (5), that is an instance of $:B$ (4); thus the subclass in (6) is a superclass of $:A$. Without (4) this entailment would be missed.

The relevant assertional information in this example was directly given, but it is clear that it could also be the result of more complicated reasoning. Therefore, it is not possible in general to compute all class subsumptions of an ontology without also computing a significant amount of fact entailments as well. Theorem 3 in Section 4 below identifies a case where assertions can be ignored.

3 A Simpler OWL RL

OWL is a very rich language that provides many redundant syntactic constructs for the convenience of ontology engineers. When specifying a rule system for

$$\begin{array}{l}
\mathbf{R}_{ca1} \frac{\text{ClassAssertion}(C a)}{\text{SubClassOf}(\text{ObjectOneOf}(a) C)} \qquad \mathbf{R}_{ca2} \frac{\text{SubClassOf}(\text{ObjectOneOf}(a) C)}{\text{ClassAssertion}(C a)} \\
\mathbf{R}_{pc1} \frac{\text{ObjectPropertyAssertion}(P a b)}{\text{SubClassOf}(\text{ObjectOneOf}(a) \text{ObjectSomeValuesFrom}(P \text{ObjectOneOf}(b)))} \\
\mathbf{R}_{pc2} \frac{\text{SubClassOf}(\text{ObjectOneOf}(a) \text{ObjectSomeValuesFrom}(P \text{ObjectOneOf}(b)))}{\text{ObjectPropertyAssertion}(P a b)}
\end{array}$$

Fig. 1. Rules for expressing assertions as class subsumptions

OWL RL, this abundance of syntax precludes a concise presentation – the W3C calculus already counts 78 rules. To avoid this problem, we introduce various simplification rules that allow us to restrict to a much smaller number of features.

Syntactic simplifications can affect the semantics of an ontology. On the one hand, they may introduce auxiliary vocabulary symbols that had not been defined by the original ontology. On the other hand, any syntactic transformation of expressions has an impact on the RDF-based Semantics of OWL, which entails only axioms about expressions that are syntactically present in the ontology. Adding a new expression, even if tautological, thus changes entailments. However, we expect applications that rely on the RDF-based Semantics to tolerate this effect. Alternatively, it is possible to view our syntactic simplifications as a mere abbreviation scheme for a much larger number of rules.

Lists in Axioms We generally assume that all lists of classes or properties in OWL axioms have been binarised, that is, broken down into lists of length two. This is always possible by introducing additional axioms; we omit the details of this frequently used technique. We point out that this simplification is least essential for our rules. It is easy to generalise all rules we state for binary lists to lists of arbitrary length.

Datatypes and Data Properties We omit all features related to datatypes from our presentation. It is not hard to add them. In essence, datatypes in OWL RL behave like classes for which certain subsumptions are given upfront (e.g., *decimal* is a superclass of *integer*), and for which some disjointness axioms are known (e.g., *rational* is disjoint from *string*). Likewise, datatype literals behave like individual names for which the membership in some class (i.e., datatype) is known. This auxiliary information can be added when loading an ontology. Thereafter, datatypes can be treated like classes, using the same rule system.

Assertions as Class Inclusions Class and property assertions can be expressed as class inclusion axioms, and indeed many rules for assertions are just special forms of the rules needed for terminological axioms. To avoid unnecessary repetition, we generally use the respective forms interchangeably. This is formalised by the rules in Fig. 1. The rules are applied top to bottom: whenever the axioms above the line are found, the axioms below the line are added. C denotes a class

Table 1. Syntactic simplifications for OWL RL constructs

ObjectMaxCardinality(0 P)	ObjectAllValuesFrom(P owl:Nothing)
ObjectMaxCardinality(0 P C)	ObjectAllValuesFrom(P ObjectComplementOf(C))
ObjectHasValue(P d)	ObjectSomeValuesFrom(P ObjectOneOf(c))
ObjectOneOf($a_1 \dots a_n$)	ObjectUnionOn(ObjectOneOf(a_1) ... ObjectOneOf(a_n))
EquivalentClasses($C_1 \dots C_n$)	SubClassOf(C_1 C_2), ..., SubClassOf(C_n C_1)
DisjointClasses($C_1 \dots C_n$)	SubClassOf(ObjectIntersectionOf(C_i C_j) owl:Nothing) for all $1 \leq i < j \leq n$
ObjectPropertyDomain(P C)	SubClassOf(ObjectSomeValuesFrom(P owl:Thing) C)
ObjectPropertyRange(P C)	SubClassOf(owl:Thing ObjectAllValuesFrom(P C))
EquivalentObjectProperties($P_1 \dots P_n$)	SubObjectPropertyOf(P_1 P_2), ..., SubObjectPropertyOf(P_n P_1)
InverseObjectProperties(P Q)	SubObjectPropertyOf(P ObjectInverseOf(Q)), SubObjectPropertyOf(Q ObjectInverseOf(P))
SymmetricObjectProperty(P)	SubObjectPropertyOf(P ObjectInverseOf(P))
TransitiveObjectProperty(P)	SubObjectPropertyOf(ObjectPropertyChain(P P) P)
FunctionalObjectProperty(P)	SubClassOf(owl:Thing ObjectMaxCardinality(1 P))
InverseFunctionalObjectProperty(P)	SubClassOf(owl:Thing ObjectMaxCardinality(1 ObjectInverseOf(P)))
AsymmetricObjectProperty(P)	DisjointObjectProperties(P ObjectInverseOf(P))
SameIndividual($a_1 \dots a_n$)	SubClassOf(ObjectOneOf(a_1) ObjectOneOf(a_2)), ..., SubClassOf(ObjectOneOf(a_n) ObjectOneOf(a_1))
NegativeObjectPropertyAssertion(P a b)	SubClassOf(ObjectOneOf(a) ObjectComplementOf(ObjectSomeValuesFrom(P ObjectOneOf(b)))
DifferentIndividuals($a_1 \dots a_n$)	SubClassOf(ObjectOneOf(a_i) ObjectComplementOf(ObjectOneOf(a_j))) for all $1 \leq i < j \leq n$

expression (not necessarily a class name), and P denotes an object property expression. We use a and b for individual names (IRIs).

We will make use of the shorter syntactic form of assertions whenever possible. Note that the class subsumptions in rules **R_{pc1}** and **R_{pc2}** are not in OWL RL, which does not allow **ObjectOneOf** in superclasses. Allowing this does not increase reasoning complexity as long as one restricts to exactly one individual in **ObjectOneOf** (indeed, this is also done in the tractable OWL EL profile). We will therefore introduce such expressions whenever this simplifies presentation.

Syntactic Sugar Many OWL features are directly expressible in terms of others. Table 1 lists straightforward syntactic transformations. C , D , E denote class expressions, P , Q object property expressions, and all lower-case letters denote individuals. Below, we can therefore disregard all features on the left of this table. As before, some of the expressions that we use here are not in OWL RL. Besides **ObjectOneOf** superclasses as discussed above, OWL RL also disallows **owl:Thing** to be used as a subclass. Again, introducing this does not complicate reasoning. The main motivation for leaving out **owl:Thing** in the standard is that it may lead to a big number of “uninteresting” entailments, since every individual is an instance of **owl:Thing**.

In summary, we therefore consider only OWL class inclusion axioms of the form **SubClassOf(CL CR)**, where **CL** and **CR** are defined as in Table 2.

Table 2. Subclasses (**CL**) and superclasses (**CR**) in syntactically simplified OWL RL

CL ::= **Class** | ObjectIntersectionOf(**CL** **CL**) | ObjectUnionOf(**CL** **CL**) |
 ObjectOneOf(**Individual**) | ObjectSomeValuesFrom(**Property** **CL**)
CR ::= **Class** | ObjectIntersectionOf(**CR** **CR**) | ObjectComplementOf(**CL**) |
 ObjectAllValuesFrom(**Property** **CR**) | ObjectMaxCardinality(1 **Property** **CL**) |
 ObjectSomeValuesFrom(**Property** ObjectOneOf(**Individual**)) | ObjectOneOf(**Individual**)

$$\begin{array}{l}
 \mathbf{C}_{\text{sco}} \quad \frac{\text{SubClassOf}(C D) \quad \text{SubClassOf}(D E)}{\text{SubClassOf}(C E)} \\
 \mathbf{C}_{\text{init}} \quad \frac{C \text{ a class expression in the ontology}}{\text{SubClassOf}(C C) \quad \text{SubClassOf}(C \text{ owl:Thing })} \\
 \mathbf{C}_{\text{int-}} \quad \frac{\text{SubClassOf}(C \text{ ObjectIntersectionOf}(D_1 D_2))}{\text{SubClassOf}(C D_1) \quad \text{SubClassOf}(C D_2)} \\
 \mathbf{C}_{\text{int+}} \quad \frac{\text{SubClassOf}(C D_1) \quad \text{SubClassOf}(C D_2)}{\text{SubClassOf}(C \text{ ObjectIntersectionOf}(D_1 D_2))} \\
 \mathbf{C}_{\text{com-}} \quad \frac{\text{SubClassOf}(C D) \quad \text{SubClassOf}(C \text{ ObjectComplementOf}(D))}{\text{SubClassOf}(C \text{ owl:Nothing })} \\
 \mathbf{C}_{\text{uni+}} \quad \frac{\text{SubClassOf}(C D) \quad \text{where } D = D_1 \text{ or } D = D_2}{\text{SubClassOf}(C \text{ ObjectUnionOf}(D_1 D_2))} \\
 \mathbf{C}_{\text{sa}} \quad \frac{\text{SubClassOf}(\text{ObjectOneOf}(c) \text{ ObjectOneOf}(d))}{\text{SubClassOf}(\text{ObjectOneOf}(d) \text{ ObjectOneOf}(c))}
 \end{array}$$

Fig. 2. OWL RL inference rules for class subsumptions

4 A Rule-Based Classification Calculus for OWL RL

In this section, we specify a class subsumption algorithm for OWL RL, and we introduce conditions under which it is complete. Using the simplifications of the previous section, the only OWL axioms that we need to consider are `SubClassOf`, `SubObjectPropertyOf`, `DisjointObjectProperties`, `IrreflexiveObjectProperty`, and `HasKey`. Remaining expressive features are those given in Table 2 (for class expressions) and `ObjectPropertyChain` (for property inclusions).

Figures 2 and 3 specify a rule system for deriving class subsumptions, where we use the same notation for rules as above. The rules in Fig. 2 apply to class subsumptions and, using the correspondences of Fig. 1, also to assertions. In contrast, the rules in Fig. 3 are only applicable to specific assertions and are not generalised to other subsumptions. For example, rule $\mathbf{P}_{\text{inv-}}$ is sound, but the following generalisation to class inclusions would of course be wrong:

$$\frac{\text{SubClassOf}(C \text{ ObjectSomeValuesFrom}(\text{ObjectInverseOf}(P) D))}{\text{SubClassOf}(D \text{ ObjectSomeValuesFrom}(P C))}$$

As an additional condition for applying rules, we require that all class and property expressions in the conclusion also occur in the ontology. This is a restriction only for the rules $\mathbf{C}_{\text{int+}}$, $\mathbf{C}_{\text{uni+}}$, $\mathbf{P}_{\text{svf+}}$, and $\mathbf{P}_{\text{inv+}}$, since they are the only rules

$$\begin{array}{l}
\mathbf{P}_{\text{avf}-} \frac{\text{ObjectPropertyAssertion}(P c d) \quad \text{ClassAssertion}(\text{ObjectAllValuesFrom}(P E) c)}{\text{ClassAssertion}(E d)} \\
\mathbf{P}_{\text{svf}+} \frac{\text{ObjectPropertyAssertion}(P c d) \quad \text{ClassAssertion}(E d)}{\text{ClassAssertion}(\text{ObjectSomeValuesFrom}(P E) c)} \\
\mathbf{P}_{\text{inv}-} \frac{\text{ObjectPropertyAssertion}(\text{ObjectInverseOf}(P) c d)}{\text{ObjectPropertyAssertion}(P d c)} \\
\mathbf{P}_{\text{inv}+} \frac{\text{ObjectPropertyAssertion}(P d c)}{\text{ObjectPropertyAssertion}(\text{ObjectInverseOf}(P) c d)} \\
\mathbf{P}_{\text{spo}} \frac{\text{ObjectPropertyAssertion}(P c d) \quad \text{SubObjectPropertyOf}(P Q)}{\text{ObjectPropertyAssertion}(Q c d)} \\
\mathbf{P}_{\text{spc}} \frac{\text{SubObjectPropertyOf}(\text{ObjectPropertyChain}(P Q) R) \quad \text{ObjectPropertyAssertion}(P c d) \quad \text{ObjectPropertyAssertion}(Q d e)}{\text{ObjectPropertyAssertion}(R c e)} \\
\mathbf{P}_{\text{dp}} \frac{\text{DisjointObjectProperties}(P Q) \quad \text{ObjectPropertyAssertion}(P c d) \quad \text{ObjectPropertyAssertion}(Q c d)}{\text{ClassAssertion}(\text{owl:Nothing } c)} \\
\mathbf{P}_{\text{ip}} \frac{\text{ObjectPropertyAssertion}(P c c) \quad \text{IrreflexiveObjectProperty}(P)}{\text{ClassAssertion}(\text{owl:Nothing } c)} \\
\mathbf{P}_{\text{key}} \frac{\text{HasKey}(E (P_1 \dots P_n) ()) \quad \text{ClassAssertion}(E c) \quad \text{ClassAssertion}(E d) \quad \text{ObjectPropertyAssertion}(P_1 c e_1) \dots \text{ObjectPropertyAssertion}(P_n c e_n) \quad \text{ObjectPropertyAssertion}(P_1 d e_1) \dots \text{ObjectPropertyAssertion}(P_n d e_n)}{\text{SubClassOf}(\text{ObjectOneOf}(c) \text{ObjectOneOf}(d))} \\
\mathbf{P}_{\text{fun}} \frac{\text{ClassAssertion}(\text{ObjectMaxCardinality}(1 P D) c) \quad \text{ObjectPropertyAssertion}(P c e_1) \quad \text{ObjectPropertyAssertion}(P c e_2) \quad \text{ClassAssertion}(D e_1) \quad \text{ClassAssertion}(D e_2)}{\text{SubClassOf}(\text{ObjectOneOf}(e_1) \text{ObjectOneOf}(e_2))}
\end{array}$$

Fig. 3. OWL RL inference rules that are specific to property assertions

that derive expressions that are not mentioned in their premise. Due to this restriction, it is easy to see that the following holds:

Lemma 1. *Every class or property expression that occurs in an axiom that is derived by applying the rules exhaustively to an ontology \mathcal{O} occurs also in \mathcal{O} .*

For an OWL ontology \mathcal{O} , we say that a class subsumption $\text{SubClassOf}(C D)$ is *inferred* by the rule system if one of the following axioms is derived by applying the rules exhaustively to \mathcal{O} :

- $\text{SubClassOf}(C D)$,
- $\text{SubClassOf}(C \text{owl:Nothing })$, or
- $\text{ClassAssertion}(\text{owl:Nothing } c)$ for some individual c .

The first condition corresponds to a direct derivation, the second captures the case that C is inconsistent (necessarily empty), and the third case occurs when-

ever \mathcal{O} is inconsistent. The inference rules in [16] use a special conclusion *false* to encode ontology inconsistency, but this just a minor difference.

Theorem 1. *The rule system is sound, that is, if a class subsumption A is inferred from \mathcal{O} , then \mathcal{O} entails A under the Direct Semantics and under the RDF-based Semantics of OWL.*

Proof. The result follows by an easy induction that shows that every class subsumption that is derived by the rule system is a logical consequence under either semantics. Clearly, this is true for the initial set of derivations, which is just the input ontology. The induction step is established by verifying that every rule is sound, that is: whenever its preconditions of the rule are satisfied in a model, then its consequences are also satisfied in this model. This is easy to see for either semantics of OWL. \square

However, the rule system is not complete. The following examples illustrate two interesting cases that are not covered.

Example 3. We consider an ontology of four axioms:

$$\text{ClassAssertion}(\text{:D} \text{:d}) \quad (7)$$

$$\text{SubClassOf}(\text{:D} \text{:C}) \quad (8)$$

$$\text{SubClassOf}(\text{:C} \text{ObjectHasValue}(\text{:P} \text{:a})) \quad (9)$$

$$\text{ClassAssertion}(\text{ObjectMaxCardinality}(1 \text{ObjectInverseOf}(\text{:P}) \text{owl:Thing}) \text{:a}) \quad (10)$$

From this, the axiom $\text{SubClassOf}(\text{:C} \text{:D})$ follows. Indeed, (9) and (10) together imply that :C can have at most one instance; by (7) and (8), this instance is :d and thus contained in :D . However, this is not entailed by our rule system. Axioms (9) and (10) can be represented as follows:

$$\text{SubClassOf}(\text{:C} \text{ObjectSomeValuesFrom}(\text{:P} \text{ObjectOneOf}(\text{:a}))) \quad (11)$$

$$\text{ClassAssertion}(\text{ObjectAllValuesFrom}(\text{ObjectInverseOf}(\text{:P}) \text{ObjectOneOf}(\text{:e})) \text{:a}) \quad (12)$$

where :e is an auxiliary individual. Using \mathbf{C}_{sco} and the rules of Fig. 1, we can derive $\text{ObjectPropertyAssertion}(\text{:P} \text{:d} \text{:a})$ from (7), (8), and (11). By applying rule $\mathbf{P}_{\text{inv}+}$, we obtain $\text{ObjectPropertyAssertion}(\text{ObjectInverseOf}(\text{:P}) \text{:a} \text{:d})$. Together with (12), $\mathbf{P}_{\text{avf}-}$ implies $\text{ClassAssertion}(\text{ObjectOneOf}(\text{:e}) \text{:d})$. The same could be derived for any other instance $\text{:d}'$ of :C , showing that all such instances must be equal, and instances of :D . However, we cannot derive $\text{SubClassOf}(\text{:C} \text{:D})$.

Example 4. Consider the ontology of the following axioms:

$$\text{SubClassOf}(\text{:C} \text{ObjectHasValue}(\text{:Q} \text{:b})) \quad (13)$$

$$\text{ObjectPropertyRange}(\text{:Q} \text{:D}) \quad (14)$$

$$\text{ObjectPropertyDomain}(\text{:Q} \text{:D}) \quad (15)$$

$$\text{SubClassOf}(\text{:D} \text{ObjectHasValue}(\text{:P} \text{:a})) \quad (16)$$

$$\text{SubObjectPropertyOf}(\text{ObjectPropertyChain}(\text{:P} \text{ObjectInverseOf}(\text{:P})) \text{:R}) \quad (17)$$

$$\text{SubClassOf}(\text{ObjectSomeValuesFrom}(\text{:R} \text{owl:Thing}) \text{:E}) \quad (18)$$

These axioms imply $\text{SubClassOf}(\text{:C}:\text{E})$. Indeed, axioms (16) and (17) together imply that every pair of instances of :D is connected by property :R . Axioms (14) and (15) in turn imply that :Q only connects instances that are in :D . Thus, we find that :Q is a subproperty of :R , which is an interesting inference in its own right. Combining this with (13) and (18), we obtain the claimed entailment $\text{SubClassOf}(\text{:C}:\text{E})$. Again, this is not inferred by our inference rules.

Examples 3 and 4 illustrate two very different forms of semantic interactions that lead to entailments not inferred by our rule system. It is interesting to note, however, that ObjectHasValue plays a crucial role in both cases. Indeed, we find that this feature is involved in every situation where an entailment is missed:

Theorem 2. *If \mathcal{O} is an OWL RL ontology that does not contain ObjectHasValue in superclasses, then the rule system is complete for \mathcal{O} . More precisely, let \mathcal{O}' be the syntactic simplification of \mathcal{O} as discussed above. If :A and :B are class names and \mathcal{O} entails $\text{SubClassOf}(\text{:A}:\text{B})$ under Direct Semantics, then $\text{SubClassOf}(\text{:A}:\text{B})$ is inferred from \mathcal{O}' by the rule system.*

An essential ingredient for proving this result is the following observation.

Lemma 2. *Consider an OWL RL ontology \mathcal{O} that does not contain ObjectHasValue in superclasses, and let \mathcal{O}' be the syntactic simplification of \mathcal{O} .*

- (1) *If an axiom of the form $\text{SubClassOf}(C \text{ ObjectOneOf}(d))$ is derived from \mathcal{O}' , then C has the form $\text{ObjectOneOf}(c)$.*
- (2) *If an axiom of the form $\text{SubClassOf}(C \text{ ObjectSomeValuesFrom}(P D))$ is derived from \mathcal{O}' , then C has the form $\text{ObjectOneOf}(c)$, and there is an individual d such that $\text{ClassAssertion}(D d)$ and $\text{ObjectPropertyAssertion}(P c d)$ are derived.*

Proof. The properties are verified by a simple induction. All class inclusion axioms of the initial ontology \mathcal{O}' satisfy the claims, which can be seen as follows. OWL RL ontologies cannot contain ObjectOneOf in superclasses (including property domains). Axioms as in (1) can only be introduced during the syntactic simplification of SameIndividual and ObjectHasValue . The former satisfy the conditions; the latter was excluded. Axioms of the form (2) that are introduced by syntactic simplification of $\text{ObjectPropertyAssertion}$ clearly satisfy the conditions. Without ObjectHasValue in superclasses, this is the only source of non-negated (i.e., superclass) uses of $\text{ObjectSomeValuesFrom}$ in the initial ontology.

For the induction step, it remains to check that the claim holds for every axiom of the form (1) or (2) that is derived by any of the rules, under the assumption that the claim is true for all premises of the rules. Rule \mathbf{C}_{sco} can derive axioms of type (1) and (2). In either case, the second premise must also be of this form, so in particular D has the form $\text{ObjectOneOf}(d)$. Thus the first premise is of form (1), so by the induction hypothesis C has the form $\text{ObjectOneOf}(c)$. For case (2), we additionally find that axioms $\text{ObjectPropertyAssertion}(P d e)$ and $\text{ClassAssertion}(E e)$ are derived. By the rules in Fig. 1, we also obtain $\text{SubClassOf}(\text{ObjectOneOf}(d) \text{ ObjectSomeValuesFrom}(P \text{ ObjectOneOf}(e)))$. By \mathbf{C}_{sco}

we derive $\text{SubClassOf}(\text{ObjectOneOf}(c) \text{ObjectSomeValuesFrom}(P \text{ObjectOneOf}(e)))$, which leads to the required $\text{ObjectPropertyAssertion}(P c e)$, again by the rules of Fig. 1.

Further rules that can derive axioms of type (1) are: \mathbf{C}_{init} applied to expressions $\text{ObjectOneOf}(c)$, \mathbf{C}_{sa} , \mathbf{P}_{key} , and \mathbf{P}_{fun} . The rules that can derive axioms of type (2) are: $\mathbf{P}_{\text{svf+}}$, $\mathbf{P}_{\text{inv-}}$, $\mathbf{P}_{\text{inv+}}$, \mathbf{P}_{spo} , \mathbf{P}_{spc} . All other rules cannot derive axioms of the forms that we consider. For rules $\mathbf{C}_{\text{int-}}$, $\mathbf{C}_{\text{int+}}$, $\mathbf{C}_{\text{com-}}$, $\mathbf{C}_{\text{uni+}}$, and $\mathbf{P}_{\text{avf-}}$ this is a consequence of Lemma 1. Most derivations immediately satisfy the claim since they are stated using individuals. The only exception is $\mathbf{P}_{\text{svf+}}$, but this rule has the claim as its preconditions. \square

We are now ready to show the claimed completeness of the calculus.

Proof (of Theorem 2). Let \mathcal{O} be the syntactic simplification of an OWL RL ontology that does not contain ObjectHasValue , and let \mathcal{O}' be the saturation of \mathcal{O} under the above rules. If \mathcal{O}' contains an axiom of the form $\text{SubClassOf}(c \text{owl:Nothing})$, then the claim follows immediately since in this case the calculus is defined to infer every class inclusion axiom.

For the remainder, we assume that \mathcal{O}' does not contain an axiom of the form $\text{SubClassOf}(\text{ObjectOneOf}(c) \text{owl:Nothing})$. We define an interpretation \mathcal{I} as follows: The domain $\Delta^{\mathcal{I}}$ of \mathcal{I} contains

- all individual names (without loss of generality, we assume that there is at least one, even if it does not occur in \mathcal{O}), and
- an element of the form e_C for every class expression C in \mathcal{O} such that $\text{SubClassOf}(C \text{owl:Nothing}) \notin \mathcal{O}'$,

where we identify every pair of individual names c and d for which $\text{SubClassOf}(\text{ObjectOneOf}(c) \text{ObjectOneOf}(d)) \in \mathcal{O}'$. All other elements are mutually distinct. We assign a class expression $\text{cls}(d)$ to every element $d \in \Delta^{\mathcal{I}}$: if d is an individual name, then $\text{cls}(d) := \text{ObjectOneOf}(d)$; if $d = e_C$ then $\text{cls}(d) := C$. The interpretation \mathcal{I} is further defined as follows:

- For every individual name c , define $c^{\mathcal{I}} := c$.
- For every class name A , define $d \in A^{\mathcal{I}}$ if and only if $\text{SubClassOf}(\text{cls}(d) A) \in \mathcal{O}'$.
- For every property name P , define $\langle c, d \rangle \in P^{\mathcal{I}}$ if and only if c and d are individual names and $\text{ObjectPropertyAssertion}(P c d) \in \mathcal{O}'$.

It is easy to see that \mathcal{I} is well-defined. In particular, individual names c and d that are identified will be assigned to the same class and property extensions since \mathcal{O}' is saturated under \mathbf{C}_{sub} , \mathbf{C}_{sa} , and $\mathbf{P}_{\text{svf+}}$. We show that \mathcal{I} is a model of \mathcal{O} that satisfies only those class subsumptions between class names that occur in \mathcal{O}' . We first observe that the relationship used to define \mathcal{I} for property names extends to inverse properties:

Claim 1 The following statements are equivalent:

- $\text{ObjectInverseOf}(P)$ occurs in \mathcal{O} and $\langle c, d \rangle \in \text{ObjectInverseOf}(P)^{\mathcal{I}}$
- c and d are individual names and $\text{ObjectPropertyAssertion}(\text{ObjectInverseOf}(P) \ c \ d) \in \mathcal{O}'$

If the first statement holds, then $\langle d, c \rangle \in P^{\mathcal{I}}$ by the semantics of inverse properties. By definition of \mathcal{I} , c and d are individual names, and $\text{ObjectPropertyAssertion}(P \ d \ c) \in \mathcal{O}'$. Thus rule $\mathbf{P}_{\text{inv}+}$ is applicable. Since \mathcal{O}' is saturated, we find that $\text{ObjectPropertyAssertion}(\text{ObjectInverseOf}(P) \ c \ d) \in \mathcal{O}'$. Conversely, if the second statement holds, then $\mathbf{P}_{\text{inv}-}$ is applicable. Thus $\text{ObjectPropertyAssertion}(P \ d \ c) \in \mathcal{O}'$. By the definition of \mathcal{I} , $\langle d, c \rangle \in P^{\mathcal{I}}$. Therefore $\langle c, d \rangle \in \text{ObjectInverseOf}(P)^{\mathcal{I}}$, as required. The fact that $\text{ObjectInverseOf}(P)$ occurs in \mathcal{O} follows by Lemma 1. This completes the proof of Claim 1.

Claim 2 If $E \in \mathbf{CL}$ occurs in \mathcal{O} , then $c \in E^{\mathcal{I}}$ implies $\text{SubClassOf}(\text{cls}(c) \ E) \in \mathcal{O}'$.

In other words, all relationships $c \in E^{\mathcal{I}}$ that hold in \mathcal{I} are also derived in \mathcal{O}' . To show this claim for arbitrary \mathbf{CL} classes E , we perform an induction over the structure of such classes, as defined by the grammar in Table 2. We begin with the base cases:

- If E is a class name, then the claim follows from the definition of $E^{\mathcal{I}}$.
- If $E = \text{owl:Nothing}$, then the claim holds since $c \in E^{\mathcal{I}}$ does not hold for any c (since $\text{owl:Nothing}^{\mathcal{I}} = \emptyset$).
- If $E = \text{owl:Thing}$, then the claim follows from rule \mathbf{C}_{init} .

For the remaining cases, we assume that the claim has already been established for the classes D , D_1 and D_2 (induction hypothesis). In all cases, we assume that E occurs in \mathcal{O} and that $c \in E^{\mathcal{I}}$.

- Case $E = \text{ObjectIntersectionOf}(D_1 \ D_2)$. According to the semantics of $\text{ObjectIntersectionOf}$, we find $c \in D_1^{\mathcal{I}}$ and $c \in D_2^{\mathcal{I}}$. Clearly, D_1 and D_2 occur in \mathcal{O} since E does. Thus, the induction hypothesis implies $\text{SubClassOf}(\text{cls}(c) \ D_1) \in \mathcal{O}'$ and $\text{SubClassOf}(\text{cls}(c) \ D_2) \in \mathcal{O}'$. Since E occurs in \mathcal{O} , rule $\mathbf{C}_{\text{int}+}$ applies and $\text{SubClassOf}(\text{cls}(c) \ E) \in \mathcal{O}'$ as required.
- Case $E = \text{ObjectUnionOf}(D_1 \ D_2)$. By the semantics of ObjectUnionOf , we find $c \in D_1^{\mathcal{I}}$ or $c \in D_2^{\mathcal{I}}$. Clearly, D_1 and D_2 occur in \mathcal{O} since E does. Thus, the induction hypothesis implies $\text{SubClassOf}(\text{cls}(c) \ D_1) \in \mathcal{O}'$ or $\text{SubClassOf}(\text{cls}(c) \ D_2) \in \mathcal{O}'$. Since E occurs in \mathcal{O} , rule $\mathbf{C}_{\text{uni}+}$ applies and $\text{SubClassOf}(\text{cls}(c) \ E) \in \mathcal{O}'$ as required.
- Case $E = \text{ObjectOneOf}(d)$. Then $c = d^{\mathcal{I}}$. By definition of \mathcal{I} , c and d denote the same domain element, and in particular c is an individual name and $\text{cls}(c) = \text{ObjectOneOf}(c)$. By the definition of \mathcal{I} , we find that either $c = d$, or $\text{SubClassOf}(\text{ObjectOneOf}(c) \ \text{ObjectOneOf}(d)) \in \mathcal{O}'$ or $\text{SubClassOf}(\text{ObjectOneOf}(d) \ \text{ObjectOneOf}(c)) \in \mathcal{O}'$. The claim is immediate in the first two cases; in the last one, we get $\text{SubClassOf}(\text{ObjectOneOf}(c) \ \text{ObjectOneOf}(d)) \in \mathcal{O}'$ from \mathbf{C}_{sa} .

- Case $E = \text{ObjectSomeValuesFrom}(P D)$. By the semantics of **ObjectSomeValuesFrom**, there is an element $d \in \Delta^{\mathcal{I}}$ such that $\langle c, d \rangle \in P^{\mathcal{I}}$ and $d \in D^{\mathcal{I}}$. By the definition of \mathcal{I} , d is an individual name. Since E occurs in \mathcal{O} , so do P and D . According to Claim 1 (if P is inverse) or the definition of \mathcal{I} (otherwise), we find that $\text{ObjectPropertyAssertion}(P c d) \in \mathcal{O}'$. By the induction hypothesis, $\text{SubClassOf}(\text{cls}(d) D) \in \mathcal{O}'$, which is the same as $\text{ClassAssertion}(D d) \in \mathcal{O}'$ since d is an individual name. Since E occurs in \mathcal{O} , rule $\mathbf{P}_{\text{svf}+}$ applies and $\text{SubClassOf}(\text{cls}(c) E) \in \mathcal{O}'$ as required.

This finishes the proof of Claim 2. The other direction needs to be established for another language of class expressions:

Claim 3 If $E \in \mathbf{CR}$ and $\text{SubClassOf}(\text{cls}(c) E) \in \mathcal{O}'$, then E occurs in \mathcal{O} and $c \in E^{\mathcal{I}}$.

The fact that E occurs in \mathcal{O} is a consequence of Lemma 1. To show the remaining claim for arbitrary **CR** classes E , we perform an induction over the structure of such classes, as defined by the grammar in Table 2. We begin with the base cases:

- If E is a class name, then the claim follows from the definition of $E^{\mathcal{I}}$.
- If $E = \text{owl:Nothing}$, then we distinguish two cases. If c is an individual name, the claim holds since we assumed that no axiom of the form $\text{SubClassOf}(\text{ObjectOneOf}(c) \text{owl:Nothing})$ occurs in \mathcal{O}' . If c is of the form e_D , then $\text{cls}(c) = D$ and $\text{SubClassOf}(D \text{owl:Nothing}) \notin \mathcal{O}'$ by construction of \mathcal{I} .
- If $E = \text{owl:Thing}$, then the claim holds since $E^{\mathcal{I}} = \Delta^{\mathcal{I}}$.

For the remaining cases, we assume that the claim has already been established for the classes D , D_1 and D_2 (induction hypothesis). In all cases, we assume that $\text{SubClassOf}(\text{cls}(c) E) \in \mathcal{O}'$.

- Case $E = \text{ObjectIntersectionOf}(D_1 D_2)$. Then rule $\mathbf{C}_{\text{int}-}$ is applicable. Since \mathcal{O}' is saturated, we find $\text{SubClassOf}(\text{cls}(c) D_1) \in \mathcal{O}'$ and $\text{SubClassOf}(\text{cls}(c) D_2) \in \mathcal{O}'$. By the induction hypothesis, $c \in D_1^{\mathcal{I}}$ and $c \in D_2^{\mathcal{I}}$. By the semantics of **ObjectIntersectionOf**, this implies $c \in E^{\mathcal{I}}$ as required.
- Case $E = \text{ObjectComplementOf}(D)$ where $D \in \mathbf{CL}$. As argued in case $E = \text{owl:Nothing}$ above, \mathcal{O}' does not contain $\text{SubClassOf}(\text{cls}(c) \text{owl:Nothing})$. Therefore, rule $\mathbf{C}_{\text{com}-}$ is not applicable, that is, $\text{SubClassOf}(\text{cls}(c) D) \notin \mathcal{O}'$. We already noted above that E and thus D must occur in \mathcal{O} . Therefore, (the contrapositive of) Claim 2 implies that $c \notin D^{\mathcal{I}}$. By the semantics of **ObjectComplementOf**, $c \in E^{\mathcal{I}}$.
- Case $E = \text{ObjectAllValuesFrom}(P D)$. Consider an arbitrary element d such that $\langle c, d \rangle \in P^{\mathcal{I}}$. According to Claim 1 (if P is inverse) or the definition of \mathcal{I} (otherwise), we find that $\text{ObjectPropertyAssertion}(P c d) \in \mathcal{O}'$. Therefore, rule $\mathbf{P}_{\text{avf}-}$ is applicable to obtain $\text{ClassAssertion}(D d) \in \mathcal{O}'$. By the induction hypothesis, this implies $d \in D^{\mathcal{I}}$. Since d was arbitrary, this shows that $c \in E^{\mathcal{I}}$ according to the semantics of **ObjectAllValuesFrom**.

- Case $E = \text{ObjectMaxCardinality}(1 P D)$. Consider two arbitrary elements d_1 and d_2 such that $\langle c, d_i \rangle \in P^{\mathcal{I}}$ and $d_i \in D^{\mathcal{I}}$. As in the previous case, we find that $\text{ObjectPropertyAssertion}(P c d_i) \in \mathcal{O}'$. By Claim 2, we find $\text{SubClassOf}(\text{cls}(d_i) D) \in \mathcal{O}'$, where $\text{cls}(d_i) = \text{ObjectOneOf}(d_i)$ since d_i occurs in a property extension. Therefore, rule \mathbf{P}_{fun} is applicable to obtain $\text{SubClassOf}(\text{ObjectOneOf}(d_1) \text{ObjectOneOf}(d_2)) \in \mathcal{O}'$. By the definition of \mathcal{I} , the domain elements for d_1 and d_2 are thus the same. Since d_1 and d_2 was arbitrary, this shows that $c \in E^{\mathcal{I}}$ according to the semantics of $\text{ObjectMaxCardinality}$.
- Case $E = \text{ObjectSomeValuesFrom}(P D)$ where $D = \text{ObjectOneOf}(d)$. By Lemma 2, c must be an individual name, and there is an individual name d' for which $\text{ObjectPropertyAssertion}(P c d') \in \mathcal{O}'$ and $\text{ClassAssertion}(D d') \in \mathcal{O}'$. By the definition of \mathcal{I} , the domain elements of d and d' are the same. By Claim 1, $\langle c, d \rangle \in P^{\mathcal{I}}$, even if P is an inverse property. By the induction hypothesis, $d \in D^{\mathcal{I}}$. By the semantics of $\text{ObjectSomeValuesFrom}$, we thus obtain $c \in E^{\mathcal{I}}$.
- Case $E = \text{ObjectOneOf}(d)$. By Lemma 2, c must be an individual name. By the definition of \mathcal{I} , the domain elements of c and d are the same. By the semantics of ObjectOneOf , we thus obtain $c \in E^{\mathcal{I}}$.

This finishes the proof of Claim 3. We can now show that \mathcal{I} is a model of \mathcal{O} . We need to show that \mathcal{I} satisfies each axiom in \mathcal{O} . We distinguish the possible forms of syntactically simplified OWL RL axioms:

- $\text{SubClassOf}(D E) \in \mathcal{O}$. Whenever there is an element $c \in D^{\mathcal{I}}$, then $\text{SubClassOf}(\text{cls}(c) D) \in \mathcal{O}'$ by Claim 2. Thus, rule \mathbf{C}_{sub} is applicable and yields $\text{SubClassOf}(\text{cls}(c) E) \in \mathcal{O}'$. By Claim 3, $c \in E^{\mathcal{I}}$. The same reasoning applies to assertions, using Fig. 1.
- $\text{SubObjectPropertyOf}(P Q) \in \mathcal{O}$, where P is a property expression or a property chain. If there are elements $\langle c, d \rangle \in P^{\mathcal{I}}$, then \mathcal{O} contains a corresponding (chain of) property assertion(s) according to Claim 1 and the definition of \mathcal{I} . Thus, rule \mathbf{P}_{sbo} or \mathbf{P}_{spc} is applicable to obtain $\text{ObjectPropertyAssertion}(Q c d)$. By Claim 1 and the definition of \mathcal{I} , we find $\langle c, d \rangle \in Q^{\mathcal{I}}$.
- $\text{DisjointObjectProperties}(P Q) \in \mathcal{O}$. Suppose for a contradiction that the axiom is not satisfied. Then there are individuals c and d with $\langle c, d \rangle \in P^{\mathcal{I}}$ and $\langle c, d \rangle \in Q^{\mathcal{I}}$. By Claim 1 and the definition of \mathcal{I} , $\text{ObjectPropertyAssertion}(P c d) \in \mathcal{O}'$ and $\text{ObjectPropertyAssertion}(Q c d) \in \mathcal{O}'$. Thus rule \mathbf{P}_{dp} applies and yields $\text{SubClassOf}(\text{ObjectOneOf}(c) \text{owl:Nothing})$. This contradicts our initial assumption.
- $\text{IrreflexiveObjectProperty}(P) \in \mathcal{O}$. Similar to $\text{DisjointObjectProperties}$.
- $\text{HasKey}(D (P_1 \dots P_n) ()) \in \mathcal{O}$. Similar to $\text{DisjointObjectProperties}$.

This shows that \mathcal{I} models all axioms of \mathcal{O} .

Finally, the overall claim can be shown by a similar argument: If a subsumption $\text{SubClassOf}(A B)$ for class names A and B is not derived from \mathcal{O} , then $\text{SubClassOf}(A B) \notin \mathcal{O}'$ and $\text{SubClassOf}(A \text{owl:Nothing}) \notin \mathcal{O}'$. Thus there is an element $e_A \in \Delta^{\mathcal{I}}$ and $\text{cls}(e_A) = A$. By definition, $e_A \in A^{\mathcal{I}}$. By the contrapositive

of Claim 2, $e_A \notin B^{\mathcal{I}}$. Hence $\mathcal{O} \not\models \text{SubClassOf}(A B)$. We have thus shown that, whenever axioms of this type are not derived, they are not logical consequences of \mathcal{O} . This finishes the proof. \square

Even in the case of Theorem 2, we need to take assertions into account during reasoning, since they might make the ontology inconsistent. For consistent ontologies, however, the rule system can be simplified further.

Theorem 3. *For a consistent OWL RL ontology \mathcal{O} that does not contain `ObjectHasValue` in superclasses, all entailed subsumptions between class names can be computed using the rules of Fig. 2 only, without taking assertions into account.*

More precisely, let $\mathcal{O}_t \subseteq \mathcal{O}$ be the set of all axioms in \mathcal{O} that do not use `ClassAssertion`, `ObjectPropertyAssertion`, `SameIndividual`, `DifferentIndividuals`, or `NegativeObjectPropertyAssertion`. Let \mathcal{O}'_t be the syntactic simplification of \mathcal{O}_t . If \mathcal{O} entails `SubClassOf(:A :B)` under Direct Semantics, and `:A` and `:B` are class names, then `SubClassOf(:A :B)` or `SubClassOf(:A owl:Nothing)` is derived from \mathcal{O}'_t by the rules of Fig. 2.

Proof. By Theorem 2, the original rule system is complete. If \mathcal{O} is consistent, we are not interested in inferences of the form `SubClassOf(ObjectOneOf(c) owl:Nothing)` but only in inferences of the form `SubClassOf(:A :B)` or `SubClassOf(:A owl:Nothing)`. We show that, whenever such inferences are obtained at all, they can be obtained by applying the rules of Fig. 2 to terminological axioms only.

To this end, we note that, whenever a rule is applied to a premise that contains individual names, the consequence must also contain individual names. For \mathbf{C}_{sa} and the rules in Fig. 3, the conclusion holds immediately. For rule \mathbf{C}_{sub} , the claim follows from Lemma 2: if D is of the form `ObjectOneOf(d)`, then C is of the form `ObjectOneOf(c)` due to the first premise. The same reasoning applies to \mathbf{C}_{com} . For the remaining rules, it is clear that all expressions in the premises occur in the conclusion as well. \square

This tells us that the computation of class subsumptions in OWL RL is indeed very simple for consistent ontologies without `ObjectHasValue` in superclasses. Provided that this situation can be assumed, it would therefore be feasible to pre-compute class subsumptions without taking assertions into account. In data-centric ontologies, this can lead to a much smaller set of axioms. In addition, the set of rules that need to be applied becomes relatively small as well.

5 RDF-Based Rule Systems

We have formulated rules above using the Functional-Style Syntax of OWL. In this section, we explain how these results transfer to RDF-based rules in the style of the W3C specification [16], which act on the RDF serialisation of OWL [18]. This also prepares the ground for discussing the limitations of such rules in the next section.

Definition 1. An RDF-based rule is a first-order implication of the form

$$T(s_1, t_1, u_1) \wedge \dots \wedge T(s_n, t_n, u_n) \rightarrow T(s, t, u)$$

where $s_{(i)}$, $t_{(i)}$, and $u_{(i)}$ are RDF terms (i.e., IRIs, literals, or blank nodes) or first-order logic variables. All variables are assumed to be universally quantified.

An (RDF-based) rule system is a finite set \mathcal{R} of RDF-based rules. \mathcal{R} is applied to an RDF graph by considering RDF triples $\langle s p o \rangle$ as facts $T(s, p, o)$. A rule system \mathcal{R} is applied to an OWL ontology \mathcal{O} by applying it to the RDF serialisation of \mathcal{O} [18].

A rule system \mathcal{R} is a sound and complete classification system for a class \mathcal{C} of ontologies if, for every ontology \mathcal{O} in \mathcal{C} and all class names $:A$ and $:B$:

$$\begin{aligned} \text{SubClassOf}(:A :B) \text{ is entailed by } \mathcal{O} \\ \text{if and only if} \\ \mathcal{R} \text{ infers } T(:A, \text{rdfs:subClassOf}, :B) \text{ from } \mathcal{O}. \end{aligned}$$

The main features of RDF-based rule systems are thus as follows:

- Finiteness: the set of rules is finite
- Monotonicity: the larger the set of (derived) facts, the larger the number of rules that can be applied
- No value invention: applying rules does not introduce new terms
- Uniformity: the applicability of rules does not depend on the IRIs of the entities it is applied to, but only on the statements that these IRIs occur in
- Triple-based: the only relation symbol used in inferences is the ternary T

The W3C OWL RL inference rules do not constitute a rule system in the sense of Definition 1, since they are not finite in number. The reason is that the rules support list-based OWL features for lists of arbitrary length, leading to an infinite number of possible patterns. A rule system that is not restricted to be finite can trivially solve all reasoning tasks: for every ontology \mathcal{O} for which we require an inference $T(s, p, o)$, we can add a rule $T_{\mathcal{O}} \rightarrow T(s, p, o)$, where $T_{\mathcal{O}}$ is the triple pattern that corresponds to the RDF serialisation of \mathcal{O} . It is also clear, however, that the infinite W3C rule system does not use this potential power.

Theorem 4. *The rules of Section 4 give rise to an RDF-based sound and complete classification system for (the syntactic simplification of) OWL RL ontologies without ObjectHasValue in superclasses.*

Proof. According to Theorems 1 and 2, the rule system of Section 4 is sound and complete. Recall that a class subsumption can be derived by this rule system in three ways, corresponding to direct inference, inconsistent subclass, and inconsistent ontology. Definition 1 requires all entailments to be produced directly. To achieve this, we add the following rules:

$$\frac{C \text{ a class expression in the ontology}}{\text{SubClassOf}(\text{owl:Nothing } C)} \quad (19)$$

$$\frac{\text{ClassAssertion}(\text{owl:Nothing } c) \quad C \text{ a class expression in the ontology}}{\text{SubClassOf}(C \text{ owl:Nothing)}} \quad (20)$$

Rule (19) together with rule \mathbf{C}_{sco} ensures that an inconsistent class (one that is a subclass of `owl:Nothing`) is a subclass of every class in the ontology. Rule (20) infers all classes to be inconsistent if the ontology is inconsistent. Other rules could be used to achieve a similar effect.²

The proof is finished by observing that rule (19), rule (20), and all rules of Section 4 can easily be expressed in the RDF serialisation of OWL based on the T predicate. For example, $\mathbf{C}_{\text{int+}}$ can be written as follows, where question marks denote variables as in [16]:

$$\begin{aligned} & T(?x, \text{rdfs:subClassOf}, ?y_1) \wedge T(?x, \text{rdfs:subClassOf}, ?y_2) \wedge \\ & T(?c, \text{owl:intersectionOf}, ?l) \wedge T(?l, \text{rdf:first}, ?y_1) \wedge T(?l, \text{rdf:rest}, ?l') \wedge \\ & \quad T(?l', \text{rdf:first}, ?y_2) \wedge T(?l', \text{rdf:rest}, \text{rdf:nil}) \\ & \rightarrow T(?x, \text{rdfs:subClassOf}, ?c) \end{aligned}$$

This is also the rule that is mainly missing from [16]. Note how the check for the existence of the expression `ObjectIntersectionOf(D1 D2)` that is required for applying $\mathbf{C}_{\text{int+}}$ is performed naturally as part of the rule application. This does not work for the rules in Fig. 1, which do not require existence of all class expressions in the consequence. Either they are created as part of the syntactic simplification stage, or the remaining rules need to be extended to allow for multiple alternative forms of each premise. The latter corresponds to the approach taken in the W3C specification.

The existence of a class expression as used in rule \mathbf{C}_{init} , (19), and (20) can be expressed using prefixes such as $T(?c, \text{rdf:type}, \text{owl:Class})$. It is straightforward to express all remaining rules. \square

6 Rule-Based Classification of Unrestricted OWL RL

So far, we have seen that the (RDF-based) rule systems for OWL RL can be extended to obtain a sound and complete approach for computing class subsumptions, as long as we disallow `ObjectHasValue` in superclasses. The natural question is whether this can be extended to arbitrary OWL RL ontologies. The answer is *no*, as we will show in this section.

This negative result is caused by the use of RDF as a basis for deduction rules. Without this restriction, it is not hard to find rule-based classification systems that are sound and complete, though not necessarily efficient. Namely, one can always check `SubClassOf(C D)` by adding an axiom `ClassAssertion(C e)` for a new individual e , and checking whether `ClassAssertion(D e)` using the existing rule-based approaches for instance retrieval. The problem is that this executes a single check, based on a modification of the ontology, rather than computing all class subsumptions at once. Tests for different classes may lead

² In practice it might be preferable to check for the original conditions separately instead of deriving potentially large amounts of further axioms. In Section 6 below, however, it is more convenient to consider only one way of deriving a result.

to different inferences. To address this, we can augment each inferred axiom with the test class C for which we have assumed $\text{ClassAssertion}(C\ e)$ (the IRI of e is immaterial and does not need to be recorded). The rules are restricted to the case that all premises can be derived under the same assumption, and computation can proceed concurrently without interactions between independent assumptions. Similar solutions and various optimisations have been suggested and implemented for OWL EL [15,11]. Unfortunately, however, it is not obvious how to represent such an extended form of inferences in RDF without introducing new entities or blank nodes.

The main result of this section is that this problem is not just due to an overly naive approach for extending the rules, but rather an inherent limitation of RDF-based rules:

Theorem 5. *There is no RDF-based sound and complete classification system for OWL RL.*

This is a very general result, since it makes a statement about *every* conceivable RDF-based rule system. A technique for proving such results has been developed in the context of rule-based reasoning for OWL EL [15]. Recalling this in full detail is beyond the scope of this paper. Instead, we extract the main insights on the level of ontological reasoning (Lemma 3), and present the crucial steps for applying this approach to our scenario.

The argumentation is based on the analysis of derivations in rule systems, which can be represented as *proof trees*. The key observation is that every proof tree of a rule system can be used to construct further proof trees by selectively renaming constants (i.e., IRIs). This renaming leads to proof trees that derive the same conclusion, applying rules in the same order, but based on a renamed input ontology. The renaming may not be uniform, that is, multiple occurrences of the same constant might be renamed differently. The proof uses the fact that such renaming may destroy entailments. Ontologies that are particularly sensitive in this respect are called *critical*:

Definition 2. *A renaming of an ontology \mathcal{O} is an ontology \mathcal{O}' that is obtained from \mathcal{O} by replacing occurrences of entity names by fresh entity names, where neither the old nor the new entities have a special semantics in OWL. A renaming is uniform if all occurrences of one entity have been replaced in the same way; otherwise it is non-uniform.*

An ontology \mathcal{O} is critical for an axiom A if A is entailed by \mathcal{O} , and A is not entailed by any non-uniform renaming of \mathcal{O} .

Roughly speaking, an ontology is critical if all of its axioms are really needed for obtaining the desired conclusion. The next result explains the significance of critical ontologies \mathcal{O} . It states that, given a sound and complete classification system, it must be possible to decompose \mathcal{O} into sets $\mathcal{O}_{\text{split}}$ and $\mathcal{O} \setminus \mathcal{O}_{\text{split}}$. The requirement is that both sets “touch” in at most 3 axioms, which reflects the arity of RDF triples. By finding a critical ontology for which this is not possible, we can show that there is no sound and complete classification system.

Lemma 3. *Suppose that \mathcal{R} is a sound and complete classification system with at most ℓ atoms $T(s, t, u)$ in the premise of any rule, and consider an ontology \mathcal{O} that is critical for an axiom $\text{SubClassOf}(C D)$ with class names C and D .*

For every $\mathcal{O}' \subseteq \mathcal{O}$ with $|\mathcal{O}'| > 3(\ell + 1)$, there is $\mathcal{O}_{\text{split}} \subseteq \mathcal{O}$ such that:

- $|\mathcal{O}' \cap \mathcal{O}_{\text{split}}| \geq 4$,
- $|\mathcal{O}' \cap (\mathcal{O} \setminus \mathcal{O}_{\text{split}})| \geq 4$,
- *there are at most 3 axioms in $\mathcal{O}_{\text{split}}$ that share any vocabulary symbols with $\mathcal{O} \setminus \mathcal{O}_{\text{split}}$.*

This result is obtained as a summary of various insights from [15]. Lemma 3 requires critical ontologies to contain at least $3(\ell + 1)$ axioms. Since ℓ depends on the rule system we consider, we require critical OWL RL ontologies of arbitrary size. The following definition provides this crucial ingredient.

Definition 3. *For every natural number $k \geq 0$, we define an ontology \mathcal{O}_k as follows. We consider class names A, B, D_i ($i = 0, \dots, k + 1$), property names V, W, P_i ($i = 0, \dots, k$), Q_i ($i = 0, \dots, k + 1$), and individual names a, b, c, d_i ($i = 1, \dots, k + 1$). The ontology \mathcal{O}_k consists of the following axioms:*

$$\text{SubClassOf}(D_i \text{ ObjectHasValue}(P_i d_{i+1})) \quad 0 \leq i \leq k \quad (21)$$

$$\text{SubClassOf}(D_i \text{ ObjectAllValuesFrom}(P_i D_{i+1})) \quad 0 \leq i \leq k \quad (22)$$

$$\text{SubClassOf}(D_0 \text{ ObjectHasValue}(W a)) \quad (23)$$

$$\text{SubClassOf}(D_0 \text{ ObjectAllValuesFrom}(W A)) \quad (24)$$

$$\text{SubClassOf}(A C) \quad (25)$$

$$\text{SubClassOf}(C \text{ ObjectHasValue}(Q_{k+1} b)) \quad (26)$$

$$\text{SubClassOf}(A \text{ ObjectHasValue}(V c)) \quad (27)$$

$$\text{SubClassOf}(D_{k+1} \text{ ObjectHasValue}(V c)) \quad (28)$$

$$\text{InverseFunctionalObjectProperty}(V) \quad (29)$$

$$\text{SubObjectPropertyOf}(\text{ObjectPropertyChain}(P_i Q_{i+1}) Q_i) \quad 0 \leq i \leq k \quad (30)$$

$$\text{SubClassOf}(\text{ObjectHasValue}(Q_0 b) B) \quad (31)$$

Lemma 4. *For every $k \geq 0$, \mathcal{O}_k entails $\text{SubClassOf}(D_0 B)$.*

This can be seen as follows: If D_0 does not contain any instances, then the statement clearly holds. If D_0 contains some instance o , then it is the start of a property chain P_0, \dots, P_k (through individuals d_1, \dots, d_{k+1}) due to (21) and (22). In particular, d_k is an instance of D_{k+1} , hence, by (28), d_k has a V successor c . Similarly, by (23) and (24), a is an instance of A . By (25), (26), and (27), a thus has a Q_{k+1} successor b and a V successor c . Since V is inverse functional (29), a must therefore be equal to d_k , so d_k has a Q_{k+1} successor b . Applying axioms (30) to the chain of d_i elements, we find that d_i has the Q_i successor b for all $1 \leq i \leq k$. Accordingly, the instance o of D_0 has Q_0 successor b . By (31), o thus is an instance of B . Since this reasoning applies to every instance o of D_0 , we find that $\text{SubClassOf}(D_0 B)$ as claimed.

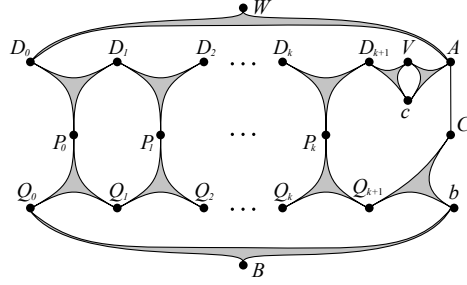


Fig. 4. Illustration of syntactic dependencies in \mathcal{O}_k

It is not hard to see that this entailment is no longer valid if any two occurrences of symbols within \mathcal{O}_k are renamed in a non-uniform way:

Lemma 5. *For every $k \geq 0$, \mathcal{O}_k is critical for $\text{SubClassOf}(D_0 B)$.*

To complete the proof of Theorem 5 we thus need to argue that there are critical ontologies that cannot be split as in Lemma 3.

Lemma 6. *Consider \mathcal{O}_k and let \mathcal{O}' be the subset of all axioms (22). For every $\mathcal{O}_{\text{split}} \subseteq \mathcal{O}$ such that $|\mathcal{O}_{\text{split}} \cap \mathcal{O}'| \geq 4$ and $|\mathcal{O}' \setminus \mathcal{O}_{\text{split}}| \geq 4$, there are at least 4 axioms in $\mathcal{O}_{\text{split}}$ that share vocabulary symbols with $\mathcal{O} \setminus \mathcal{O}_{\text{split}}$.*

Proof. We can illustrate the syntactic dependencies in an ontology by means of a (hyper)graph where each axiom is an edge between all entities that it refers to. Figure 4 shows part of the according graph for \mathcal{O}_k , showing only axioms (22), (24)–(28), (30), and (31). A subset of these axioms thus corresponds to a subset of edges, and the shared vocabulary symbols are shared nodes.

The assumptions on $\mathcal{O}_{\text{split}}$ require that $\mathcal{O}_{\text{split}}$ contains at least 4 of the axioms (22) (upper row in Fig. 4), and also misses at least 4 of the axioms (22). Using the illustration in Fig. 4, it is not hard to see that this requires $\mathcal{O}_{\text{split}}$ to share signature symbols with at least 4 axioms not in $\mathcal{O}_{\text{split}}$. \square

Thus, for any rule system \mathcal{R} with at most ℓ atoms in rule premises, the ontology \mathcal{O}_k for $k = 3(\ell + 1)$ is critical but does not satisfy the conditions of Lemma 3. Thus, \mathcal{R} cannot be a sound and complete classification system.

7 Conclusion

From a practical perspective, the main contribution of this work is to clarify the problems of OWL RL classification, and to propose rule systems for solving this task in relevant cases. The rules that we proposed produce sound conclusions on arbitrary OWL ontologies, under either of the two semantics of OWL. If the input is an OWL RL ontology where **ObjectHasValue** is not used in superclasses, the rule system is also guaranteed to be complete.

Our findings have also been brought to the attention of the OWL Working Group, which is preparing an editorial update of the OWL 2 specification at the time of this writing. This new version will correct the complexity claims about OWL RL. Extending the inference rules to be complete for computing class subsumptions in ontologies without `ObjectHasValue`, however, is beyond the scope of this editorial update. OWL RL tools can achieve completeness in this sense by adding, in essence, the one additional rule given after Theorem 4 (generalised to conjunctions of arbitrary arity). This does not affect official conformance.

Interestingly, `ObjectHasValue` in superclasses is the one OWL RL feature that complicates schema reasoning the most. This contrasts with OWL EL, where such expressions are as easy to handle as assertions [11]. The reason is that inverse properties, `ObjectAllValuesFrom`, and `ObjectMaxCardinality`, all of which allow for some complicated interactions with `ObjectHasValue`, are not in OWL EL.

Another interesting insight of this work is that there are practical problems in OWL RL reasoning that RDF-based rules are too inexpressive to solve. This limitation is due to the triple-based representation of RDF, which could be overcome by allowing predicates of higher arities as in Datalog [1] or RIF [12]. For keeping closer to features supported in RDF databases, it might be possible to use *quads* or *named graphs* for expressing 4-ary predicates, but it is doubtful if this would be an adequate use of these features. On the other hand, 4-ary relations are only needed as intermediate results during reasoning, so individual systems can implement solutions without referring to any language standard.

Another approach is to allow rules with value creation (*blank nodes* in rule heads) to encode n-ary relationships by introducing auxiliary entities. Value invention is problematic in general, as it can lead to non-termination and undecidability. Many works have studied conditions that ensure termination of bottom-up reasoning in the presence of value creation – see [4] for a recent overview – but it is unclear if any of these conditions would apply in our case.

Acknowledgements The research reported herein was supported by the EPSRC projects CONDOR and ExODA.

References

1. Abiteboul, S., Hull, R., Vianu, V.: Foundations of Databases. Addison Wesley (1994)
2. Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., Velkov, R.: OWLIM: a family of scalable semantic repositories. *Semantic Web Journal* 2(1), 33–42 (2011)
3. Brewka, G., Eiter, T., McIlraith, S.A. (eds.): Proc. 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'12). AAAI Press (2012)
4. Cuenca Grau, B., Horrocks, I., Krötzsch, M., Kupke, C., Magka, D., Motik, B., Wang, Z.: Acyclicity conditions and their application to query answering in description logics. In: Brewka et al. [3], pp. 243–253
5. Erling, O.: Virtuoso, a hybrid RDBMS/graph column store. *IEEE Data Eng. Bull.* 35(1), 3–8 (2012)

6. Franz Inc.: AllegroGraph RDFStore: Web 3.0's Database (2012), <http://www.franz.com/agraph/allegrograph/>, accessed April 2012
7. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S. (eds.): OWL 2 Web Ontology Language: Primer. W3C Recommendation (27 October 2009), available at <http://www.w3.org/TR/owl2-primer/>
8. Hogan, A., Harth, A., Polleres, A.: Scalable authoritative OWL reasoning for the Web. *Int. J. of Semantic Web Inf. Syst.* 5(2), 49–90 (2009)
9. Hogan, A., Pan, J.Z., Polleres, A., Decker, S.: SAOR: template rule optimisations for distributed reasoning over 1 billion linked data triples. In: Patel-Schneider et al. [19], pp. 337–353
10. ter Horst, H.J.: Completeness, decidability and complexity of entailment for RDF Schema and a semantic extension involving the OWL vocabulary. *J. of Web Semantics* 3(2–3), 79–115 (2005)
11. Kazakov, Y., Krötzsch, M., Simančík, F.: Practical reasoning with nominals in the \mathcal{EL} family of description logics. In: Brewka et al. [3], pp. 264–274
12. Kifer, M., Boley, H. (eds.): RIF Overview. W3C Working Group Note (22 June 2010), available at <http://www.w3.org/TR/rif-overview/>
13. Kolovski, V., Wu, Z., Eadon, G.: Optimizing enterprise-scale OWL 2 RL reasoning in a relational database system. In: Patel-Schneider et al. [19], pp. 436–452
14. Kotoulas, S., Oren, E., van Harmelen, F.: Mind the data skew: distributed inferencing by speeddating in elastic regions. In: Proc. 19th Int. Conf. on World Wide Web (WWW'10). pp. 531–540. WWW'10, ACM (2010)
15. Krötzsch, M.: Efficient rule-based inferencing for OWL EL. In: Walsh, T. (ed.) Proc. 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI'11). pp. 2668–2673. AAAI Press/IJCAI (2011)
16. Motik, B., Cuenca Grau, B., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C. (eds.): OWL 2 Web Ontology Language: Profiles. W3C Recommendation (27 October 2009), available at <http://www.w3.org/TR/owl2-profiles/>
17. Motik, B., Patel-Schneider, P.F., Parsia, B. (eds.): OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax. W3C Recommendation (27 October 2009), available at <http://www.w3.org/TR/owl2-syntax/>
18. Patel-Schneider, P.F., Motik, B. (eds.): OWL 2 Web Ontology Language: Mapping to RDF Graphs. W3C Recommendation (27 October 2009), available at <http://www.w3.org/TR/owl2-mapping-to-rdf/>
19. Patel-Schneider, P.F., Pan, Y., Glimm, B., Hitzler, P., Mika, P., Pan, J., Horrocks, I. (eds.): Proc. 9th Int. Semantic Web Conf. (ISWC'10), LNCS, vol. 6496. Springer (2010)
20. Schenk, S., Staab, S.: Networked graphs: a declarative mechanism for SPARQL rules, SPARQL views and RDF data integration on the Web. In: Huai, J., Chen, R., Hon, H.W., Liu, Y., Ma, W.Y., Tomkins, A., Zhang, X. (eds.) Proc. 17th Int. Conf. on World Wide Web (WWW'08). pp. 585–594. ACM (2008)
21. Urbani, J., Kotoulas, S., Maassen, J., van Harmelen, F., Bal, H.: WebPIE: a Web-scale parallel inference engine using MapReduce. *J. of Web Semantics* 10, 59–75 (2012)

Revision History

- 10 July 2012. Clarified proof of Theorem 4. Fixed some typos.
- 26 June 2012. Fixed a glitch in the treatment of functionality (ObjectMaxCardinality).
- 13 November 2012. Fixed some typos.