

Minimal Objectification and Maximal Unnesting in PSOA RuleML

Gen Zou, Harold Boley

Faculty of Computer Science, University of New Brunswick, Fredericton, Canada
{gen[DOT]zou, harold[DOT]boley}[AT]unb.ca

Abstract. The paper introduces two connected advancements of Positional-Slotted, Object-Applicative RuleML: (1) a model-theoretic semantics, realized transformationally, that directly handles atoms (i.e., predicate applications) without object identifiers (e.g., relationships as in Prolog) and (2) a transformational semantics that handles nested atomic formulas (e.g., nested frames as in Flora-2/F-logic). For (1), the model theory is extended to atoms with optional OIDs, the transformation is developed from static to dynamic objectification, and the correctness of the realization is proved. For (2), the unnesting transformation is defined to decompose nested atomic formulas into equivalent conjunctions.

1 Introduction

The relational and graph (object-centered) modeling paradigms have been widely used for knowledge representation in AI and the Semantic Web. The relational paradigm (e.g., classical logic and logic programming) is built on top of *relationships* with positional arguments, while the object-centered paradigm (e.g., RDF, N3, and F-logic) is built on top of *frames* with a globally unique Object Identifier (OID), usually typed by a class, and an unordered collection of slotted (attribute-value) arguments. To facilitate interoperability between the two paradigms, e.g. for expressing mappings between frames and relational databases in rule-based data access, combined object-relational paradigms have been studied. F-logic [1,2] and RIF-BLD [3] employ a heterogeneous approach to allow the combined use of relationships and frames. In contrast, the Web rule language PSOA RuleML [4] employs a homogeneous approach by generalizing relationships and frames to **positional-slotted object-applicative** (psoa) terms¹, which permit the application of a predicate (acting as a relation) to be [in an *oidless/oidful* dimension] without or with an OID – typed by the predicate (acting as a class) – and the predicate’s arguments to be [in an orthogonal dimension] *positional, slotted, or combined*.

PSOA RuleML allows the interchangeable use of oidless and oidful *atoms* (i.e., predicate applications) through the *objectification* transformation, which creates an OID for each oidless atom. Earlier, OIDs needed to be statically generated for

¹ We use the upper-cased “PSOA” as a qualifier for the language and the lower-cased “psoa” for its terms.

each oidless *psoa* term before applying the model-theoretic semantics [4,5]. This is inappropriate for *expressions* (i.e., function applications) since their functions cannot act as classes. It also causes overhead for an atom whose predicate in the clauses of the Knowledge Base (KB) is used only as a Prolog-like relation, in particular does not occur with an OID or slots (the latter also requires an OID for slottribution, explained in Section 2). To address these issues, the model theory is extended so that atoms and expressions can be interpreted without the need for an explicit OID, and an equivalence between an oidless atom and its existential oidful form is guaranteed. A novel static/dynamic objectification transformation – which is minimal by performing as little as possible in a static manner – is introduced to realize this semantics while leaving unchanged as many of the oidless atoms as possible, allowing better use of the underlying Prolog engine in our PSOATransRun [6] implementation since version 1.0, and better interoperation with ground facts in relational databases.

Frames are often nested. PSOA RuleML’s presentation syntax (PSOA/PS) generally permits embedded atoms almost everywhere. However, the semantics of embedded atoms has not been clearly defined in the earlier version of the language. In this paper, the unnesting transformation is formally defined to decompose nested atomic formulas into equivalent conjunctions, augmenting our implementation in the PSOATransRun 1.1 release.² It recursively extracts oidful atoms from an atomic formula, e.g. another atom, leaving behind their OIDs in the “trimmed” version. Unnesting is maximal by extracting atoms not only from other atoms but also from expressions. This transformational semantics is applied to every atomic formula before the model-theoretic semantics is applied.

The rest of the paper is organized as follows. Section 2 reviews the PSOA RuleML language. Section 3 gives the new semantics with minimal objectification and discusses different objectification transformations. Section 4 discusses the syntax of embedded atoms as well as the unnesting transformation. Section 5 concludes the paper.

2 PSOA RuleML

In this section we introduce the basics of the object-relational Web rule language PSOA RuleML [4], which generalizes RIF-BLD [3] and POSL [7] by introducing *psoa* terms of the following general forms:

$$\text{Oidless: } f([t_{1,1} \dots t_{1,n_1}] \dots [t_{m,1} \dots t_{m,n_m}] p_1 \rightarrow v_1 \dots p_k \rightarrow v_k) \quad (1)$$

$$\text{Oidful: } o\#f([t_{1,1} \dots t_{1,n_1}] \dots [t_{m,1} \dots t_{m,n_m}] p_1 \rightarrow v_1 \dots p_k \rightarrow v_k) \quad (2)$$

Both (1) and (2) apply a function or predicate f , possibly identified by an OID o through a membership $o\#f$ of o in f (acting as a class), to a bag of tupled arguments $[t_{i,1} \dots t_{i,n_i}]$, $i = 1, \dots, m$, and to a bag of slotted arguments $p_j \rightarrow v_j$, $j = 1, \dots, k$, representing attribute-value pairs. For the most often used special case of single-tuple *psoa* terms ($m=1$), the square brackets enclosing the tuple may be omitted.

² Available from <http://psoa.ruleml.org/transrun/1.1/local/>.

A psOA term can be interpreted as a psOA expression or a psOA atom, depending on whether f is a function or predicate. The interpretation as an expression was earlier allowed for both oidless and oidful psOA terms (i.e., (1) and (2) above) but will be restricted to oidless psOA terms here, as explained in Section 4. The interpretation as an atom is permitted for both (1) and (2) on the top-level, while restricted to (2) when embedded, as explained in Section 4.

The OID, tuples, and slots in a psOA atom are all optional. For an oidless psOA atom, without a ‘user’ OID, *objectification* can introduce a ‘system’ OID to make it oidful, as detailed in Section 3. Untyped objects are notated by using the root class `Top` as their type f . An oidful psOA atom of the form (2) is equivalent to a conjunction

$$\text{And}(\text{o\#f o\#Top}(\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}) \dots \text{o\#Top}(\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}) \\ \text{o\#Top}(\mathbf{p}_1 \rightarrow \mathbf{v}_1) \dots \text{o\#Top}(\mathbf{p}_k \rightarrow \mathbf{v}_k))$$

of its class membership, its bag of object-centered tuples (*tupribution*), and its bag of object-centered slots (*slotribution*). This *distribution* (tupribution and/or slotribution) could be physically implemented on the Web. A systematics of special kinds of psOA atoms, including (oidless) *relationships* $f(\mathbf{t}_1 \dots \mathbf{t}_n)$ and (oidful) *frames* $\text{o\#f}(\mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)$, is elaborated in [5] with many examples.

The alphabet of PSOA RuleML includes a single set `Const` of individual, function, and predicate constants – to prepare functional-logic integration as, e.g., in Relfun, Hilog, and RIF – as well as a set `Var` of variables. Constants include `Top`, denoting the root class, and ‘_’-prefixed *local constants* (e.g., `_a`). Variables are ‘?’-prefixed (e.g., `?X`).

The syntax of PSOA RuleML is built on *terms*. A *simple term* is a constant or a variable. An *atomic formula* is a psOA atom in the form of terms (1) or (2), a subclass term $\mathbf{c}_1 \## \mathbf{c}_2$, an equality term $\mathbf{t}_1 = \mathbf{t}_2$, or an external term `External(f(...))`. Complex formulas can be constructed using the Horn-like subset of first-order logic (FOL), including conjunctions `And`($\tau_1 \dots \tau_n$), disjunctions `Or`($\tau_1 \dots \tau_n$) in the rule body, top-level rule implications $\tau_1 :- \tau_2$, existential quantification `Exists` $?X_1 \dots ?X_n (\tau_1)$, and top-level universal quantification `Forall` $?X_1 \dots ?X_n (\tau_1)$. A *group formula* `Group`($\tau_1 \dots \tau_n$) wraps a set of facts and rules into a KB.

The model-theoretic semantics of PSOA RuleML is defined through semantic structures [4]. A semantic structure \mathcal{I} is a tuple $\langle \mathbf{TV}, \mathbf{DTS}, \mathbf{D}, \mathbf{D}_{\text{ind}}, \mathbf{D}_{\text{func}}, \mathbf{I}_{\text{C}}, \mathbf{I}_{\text{V}}, \mathbf{I}_{\text{psoa}}, \mathbf{I}_{\text{sub}}, \mathbf{I}_{=} , \mathbf{I}_{\text{external}}, \mathbf{I}_{\text{truth}} \rangle$. Here \mathbf{D} is a non-empty set called the domain of \mathcal{I} . \mathbf{D}_{ind} and \mathbf{D}_{func} are subsets of \mathbf{D} for individual and function interpretations. \mathbf{I}_{C} and \mathbf{I}_{V} interpret constants and variables. \mathbf{I}_{psoa} interprets predicates/functions of psOA terms as semantic functions, which will be shown in Section 3.1. \mathbf{I}_{sub} , $\mathbf{I}_{=}$, and $\mathbf{I}_{\text{external}}$ interpret subclass, equality, and external terms. A generic mapping \mathbf{I} from terms to \mathbf{D} can be defined using the above interpretation mappings. $\mathbf{I}_{\text{truth}}$ maps domain elements to the set of truth values $\mathbf{TV} = \{\mathbf{t}, \mathbf{f}\}$. We will write $\mathcal{I}.\mathbf{D}$, $\mathcal{I}.\mathbf{I}_{\text{V}}$, etc., for the components of \mathcal{I} in the rest of the paper.

Truth evaluation for formulas is determined by a recursive evaluation function $\mathbf{TVal}_{\mathcal{I}}$ defined in [4]. A semantic structure \mathcal{I} is called a *model* of a KB ϕ if

$TVal_{\mathcal{I}}(\phi) = \mathbf{t}$ and \mathcal{I} conforms to all semantic restrictions (e.g., subclass and slotribution/tupribution restrictions), denoted by $\mathcal{I} \models \phi$. A PSOA KB ϕ is said to *entail* a formula ψ , denoted by $\phi \models \psi$, if for every model \mathcal{I} of ϕ , $\mathcal{I} \models \psi$ holds.

We illustrate the syntax and semantics through an example, previewing key concepts, e.g. *non-relational* and *virtual OID*.

Example 1. Consider the following KB:

```
Group (
  Forall ?Pers ?JobTitle ?Comp1 ?Comp2 (
    _transfer(?Pers ?Comp1 ?Comp2) :-
      And(_work(?Pers ?Comp1 ?JobTitle)
          _acquire(_buyer->?Comp2 _seller->?Comp1)))
    _e1#_transfer(_Tony _Rho4biz _Chi4corp _bonus->20000)
    _work(_Kate _Rho4biz "Director")
    _a1#_acquire(_buyer->_Chi4corp _seller->_Rho4biz)
  )
)
```

We will query this KB as follows, without (left) and with (right) an OID:

```
_work(?P ?C ?J)           ?0#_work(?P ?C ?J)
_transfer(?P ?C1 ?C2)     ?0#_transfer(?P ?C1 ?C2)
```

The `_work` fact is relational while the `_acquire` fact has two slots centered on the object `_a1`. The `_transfer` rule uses premises satisfied by these facts. Its `_acquire` premise needs to be objectified to `?1#_acquire(_buyer->?Comp2 _seller->?Comp1)`, so that the generated fresh OID variable `?1` unifies with `_a1`. The `_transfer` fact is non-relational, centered on the graph-node-like object `_e1` – typed by `_transfer` acting as a class – and having one tuple of arguments as well as a slot for an optional `_bonus`. Since `_transfer` acts as a non-relational predicate in this KB atom, it is a non-relational predicate in the entire KB (although it acts as a relation in the rule).

Let us query the KB: The query `_work(?P ?C ?J)` uses the relational KB predicate `_work` in a relationship, while the query `?0#_work(?P ?C ?J)` uses `_work` non-rotationally with an OID variable `?0`, which can be bound to a virtual OID created by objectification. The `_transfer(?P ?C1 ?C2)` query uses the non-relational KB predicate `_transfer` without an OID. The `_transfer` rule is invoked, binding `?P` to `_Kate`, `?C1` to `_Rho4biz`, and `?C2` to `_Chi4corp`. The second solution is dependent on query objectification, introducing the existential OID variable `?1`, yielding `Exists ?1 (?1#_transfer(?P ?C1 ?C2))`; the `_transfer` fact is retrieved with the `_bonus` slot ignored, binding `?P` to `_Tony`, `?C1` to `_Rho4biz`, and `?C2` to `_Chi4corp` (the `?1` binding to `_e1` is not shown because of its `Exists` encapsulation). In the query `?0#_transfer(?P ?C1 ?C2)`, `_transfer` acts as a class – i.e., as a non-relational predicate – with an explicit OID variable `?0`. For successful query answering, the `_transfer` rule (e.g., its conclusion) requires objectification while the fact can be used directly. \square

More (e.g., geospatial) examples and the open-source implementation of PSOA are available online,³ for explorations in its operational semantics.

³ See PSOA's entry to the RuleML website: <http://psoa.ruleml.org>.

3 Minimal Objectification

In PSOA RuleML, each oidless psOA atom σ is understood as having an implicit OID, which allows the interchangeable use of oidless and oidful atoms of the respective term forms (1) and (2) in Section 2. The earlier semantics can only interpret an oidless psOA term after *static objectification*, which generates OIDs for all of the KB's oidless terms. This causes reasoning overhead for an atom whose predicate in the KB clauses is used only as a Prolog-like relation, e.g. does not occur with an OID or slots. Moreover, since it will turn out that it is inappropriate to give OIDs to expressions, the earlier semantics cannot deal with expression terms. In this section, we will thus first introduce a modified model-theoretic semantics to allow the direct interpretation of oidless psOA terms. Then we will discuss an objectification systematics for oidless atoms, including undifferentiated and differentiated static objectification transformations, as well as a novel static/dynamic transformation. Static/dynamic objectification is minimal in the sense that it generates as few explicit OIDs as possible, instead constructing virtual OIDs as query variable bindings. The current section assumes that the only allowed embedded complex terms are expressions, which can be achieved by maximal unnesting of embedded atoms – the central theme of Section 4.

3.1 New Semantics for Oidless PsOA Terms

The earlier semantics of PSOA RuleML [4] defined the interpretation of a psOA term as follows, where I is a generic mapping for interpreting any term:

$$\begin{aligned} I(\text{o}\#f([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)) = \\ I_{\text{psOA}}(I(f))(I(\text{o}), \\ \{\langle I(\mathbf{t}_{1,1}), \dots, I(\mathbf{t}_{1,n_1}) \rangle, \dots, \langle I(\mathbf{t}_{m,1}), \dots, I(\mathbf{t}_{m,n_m}) \rangle\}, \\ \{\langle I(\mathbf{p}_1), I(\mathbf{v}_1) \rangle, \dots, \langle I(\mathbf{p}_k), I(\mathbf{v}_k) \rangle\}) \end{aligned} \quad (3)$$

I first gives a domain element interpretation in D to each component of the psOA term, including: the OID o ; the predicate/function f ; each positional argument $\mathbf{t}_{i,j}$, $1 \leq i \leq m$, $1 \leq j \leq n_i$; each slot name \mathbf{p}_h and filler \mathbf{v}_h , $1 \leq h \leq k$. Then I_{psOA} maps $I(f) \in D$ to a semantic function of the general form $D_{\text{ind}} \times \text{SetOfFiniteBags}(D^*_{\text{ind}}) \times \text{SetOfFiniteBags}(D_{\text{ind}} \times D_{\text{ind}}) \rightarrow D$, which takes the following three arguments and returns an element $\mathbf{d} \in D$:

- The interpreted OID $I(\text{o}) \in D_{\text{ind}}$;
- The bag of interpreted tuples $\{\langle I(\mathbf{t}_{i,1}), \dots, I(\mathbf{t}_{i,n_i}) \rangle \mid 1 \leq i \leq m\}$
 $\in \text{SetOfFiniteBags}(D^*_{\text{ind}})$, where D^*_{ind} denotes the set of all finite-length tuples over D .
- The bag of interpreted slots $\{\langle I(\mathbf{p}_h), I(\mathbf{v}_h) \rangle \mid 1 \leq h \leq k\}$
 $\in \text{SetOfFiniteBags}(D_{\text{ind}} \times D_{\text{ind}})$.

I_{psOA} can be applied to $I(f)$ no matter f is a predicate or function symbol, where $I_{\text{psOA}}(I(f))$ must be a D_{ind} -valued semantic function if f is a function. Notice that in the above definitions, the mappings I and I_{psOA} are only defined

for oidful psoa terms and their predicates/functions, hence all psoa terms need to be objectified before applying the semantics. This has several shortcomings:

- Expressions should not have OIDs, which will be explained in Section 4.
- Atoms are required to have OIDs before they can be semantically evaluated, which creates overhead for each atom whose predicate in the KB clauses is used only as a Prolog-like relation.

In order to resolve these problems, we will redefine \mathbf{I} and \mathbf{I}_{psoa} to allow a direct interpretation of oidless psoa terms, and also incorporate the objectification virtually into the semantics using a logical equivalence. Specifically, the following changes are introduced:

1. The definition of \mathbf{I} for **oidful** psoa terms is changed to

$$\begin{aligned} \mathbf{I}(\text{o}\#\mathbf{f}([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)) = \\ \mathbf{I}_{\text{psoa}}(\mathbf{I}(\mathbf{f}))(\{\mathbf{I}(\text{o})\}, \\ \{\langle \mathbf{I}(\mathbf{t}_{1,1}), \dots, \mathbf{I}(\mathbf{t}_{1,n_1}) \rangle, \dots, \langle \mathbf{I}(\mathbf{t}_{m,1}), \dots, \mathbf{I}(\mathbf{t}_{1,n_m}) \rangle\}, \\ \{\langle \mathbf{I}(\mathbf{p}_1), \mathbf{I}(\mathbf{v}_1) \rangle, \dots, \langle \mathbf{I}(\mathbf{p}_k), \mathbf{I}(\mathbf{v}_k) \rangle\}) \end{aligned} \quad (4)$$

where the first argument of the semantic function $\mathbf{I}_{\text{psoa}}(\mathbf{I}(\mathbf{f}))$ is wrapped into a singleton set $\{\mathbf{I}(\text{o})\}$. This allows defining \mathbf{I} for **oidless** psoa terms separately, by using the empty set $\{\}$ as the first argument:

$$\begin{aligned} \mathbf{I}(\mathbf{f}([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)) = \\ \mathbf{I}_{\text{psoa}}(\mathbf{I}(\mathbf{f}))(\{\}, \\ \{\langle \mathbf{I}(\mathbf{t}_{1,1}), \dots, \mathbf{I}(\mathbf{t}_{1,n_1}) \rangle, \dots, \langle \mathbf{I}(\mathbf{t}_{m,1}), \dots, \mathbf{I}(\mathbf{t}_{1,n_m}) \rangle\}, \\ \{\langle \mathbf{I}(\mathbf{p}_1), \mathbf{I}(\mathbf{v}_1) \rangle, \dots, \langle \mathbf{I}(\mathbf{p}_k), \mathbf{I}(\mathbf{v}_k) \rangle\}) \end{aligned} \quad (5)$$

2. The definition of \mathbf{I}_{psoa} is changed to map \mathbf{D} to semantic functions of the form $\text{SetOfPhiSingletons}(\mathbf{D}_{\text{ind}}) \times \text{SetOfFiniteBags}(\mathbf{D}_{\text{ind}}^*) \times \text{SetOfFiniteBags}(\mathbf{D}_{\text{ind}} \times \mathbf{D}_{\text{ind}}) \rightarrow \mathbf{D}$, where $\text{SetOfPhiSingletons}(\mathbf{D}_{\text{ind}})$ is defined as $\{\{\}\} \cup \{\{\text{o}\} \mid \text{o} \in \mathbf{D}_{\text{ind}}\}$, whose elements contains the empty set $\{\}$ and a singleton set $\{\text{o}\}$ for each $\text{o} \in \mathbf{D}_{\text{ind}}$. With this definition, the semantic function $\mathbf{I}_{\text{psoa}}(\mathbf{I}(\mathbf{f}))$ can be correctly applied to the arguments in equations (4) and (5).
3. Define truth evaluation for oidless psoa terms used as atoms as follows.

$$\begin{aligned} TVal_{\mathcal{I}}(\mathbf{f}([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)) = \\ \mathbf{I}_{\text{truth}}(\mathbf{I}(\mathbf{f}([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)) \end{aligned} \quad (6)$$

Here, the following *objectification restriction* is required to capture the logical equivalence between an oidless atom (notice the absence of “ $\text{o}\#$ ” in front of “ $\mathbf{f}(\dots)$ ”) and its existentially objectified form:

$$TVal_{\mathcal{I}}(\mathbf{f}([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k)) = \mathbf{t}$$

if and only if

$$TVal_{\mathcal{I}}(\text{Exists } ?\text{O} (?0\#\mathbf{f}([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}] \mathbf{p}_1 \rightarrow \mathbf{v}_1 \dots \mathbf{p}_k \rightarrow \mathbf{v}_k))) = \mathbf{t}$$

where $?0$ is a fresh variable representing the postulated virtual OID of the atom. Notice that the restriction applies only to atoms – after unnesting (cf. Section 4) occurring only the top-level – but not to – embedded – expressions, which do not have a truth value. In the rest of the paper, we will call the semantics with and without objectification restriction *objectification-including* and *objectification-excluding* semantics, respectively.

Next we define the objectification transformation, which realizes the objectification-including semantics by transforming KBs and queries such that entailment can be established under the objectification-excluding semantics.

Definition 1. (*Entailment under objectification-excluding/-including semantics*) An interpretation \mathcal{I} is an objectification-excluding (resp., objectification-including) model of ϕ , written as $\mathcal{I} \models_{-o} \phi$ (resp., $\mathcal{I} \models_{+o} \phi$), iff $TVal_{\mathcal{I}}(\phi) = \mathbf{t}$ and $TVal_{\mathcal{I}}$ conforms to all other semantic restrictions [4] (e.g., slotribution/tupribution), while it does not (resp., does) guarantee the objectification restriction. A KB ϕ entails a query q under the objectification-excluding (resp. objectification-including) semantics, written as $\phi \models_{-o} q$ (resp., $\phi \models_{+o} q$), iff for every \mathcal{I} such that $\mathcal{I} \models_{-o} \phi$ (resp. $\mathcal{I} \models_{+o} \phi$), $\mathcal{I} \models_{-o} q$ (resp. $\mathcal{I} \models_{+o} q$) holds.

Since the default semantics is objectification-including, we will use the short notation \models for \models_{+o} in the rest of the paper.

Definition 2. (*Objectification Transformation*) An objectification transformation \mathbf{obj} with respect to a set of KBs Φ and a set of queries Q is a transformation $\Phi \times \Gamma \rightarrow \Gamma$, where Φ and Q are subsets of Γ , such that for every KB $\phi \in \Phi$ and query $q \in Q$, $\phi \models q$ iff $\mathbf{obj}(\phi, \phi) \models_{-o} \mathbf{obj}(\phi, q)$. For convenience, we write $\mathbf{obj}(q)$ for $\mathbf{obj}(\phi, q)$ if $\mathbf{obj}(\phi, q)$ is independent of ϕ , and $\mathbf{obj}(\phi)$ for $\mathbf{obj}(\phi, \phi)$.

The second argument of \mathbf{obj} is the input formula to be transformed, while the first argument is the corresponding KB of the formula providing necessary context information. Next, we define a schema for constructing such a transformation \mathbf{obj} for KBs and queries from a transformation defined only for atoms.

Definition 3. Given a set of KBs Φ and a set of queries Q , for any $\phi \in \Phi$, if $\mathbf{obj}(\phi, \tau)$ has been defined, with τ being any atom, then $\mathbf{obj}(\phi, \tau)$ can be extended for τ being a formula other than an atom, where (1) for a subclass or an equality, $\mathbf{obj}(\phi, \tau) = \tau$ is just a projection; (2) for other formulas, $\mathbf{obj}(\phi, \tau)$ is obtained by recursively applying \mathbf{obj} to each subformula of τ , while keeping the surrounding formula structure unchanged.

Note that $\mathbf{obj}(\phi, \tau)$ is the same as the formula obtained by replacing every atom ω in τ with $\mathbf{obj}(\phi, \omega)$; for some ω 's, this is just a projection $\mathbf{obj}(\phi, \omega) = \omega$. In the rest of the paper we assume every \mathbf{obj} is constructed using Definition 3.

In order to prove $\phi \models q$ iff $\mathbf{obj}(\phi) \models_{-o} \mathbf{obj}(\phi, q)$, as required by Definition 2 for any particular \mathbf{obj} , one can prove $\phi \models q$ iff $\mathbf{obj}(\phi) \models \mathbf{obj}(\phi, q)$ and $\mathbf{obj}(\phi) \models \mathbf{obj}(\phi, q)$ iff $\mathbf{obj}(\phi) \models_{-o} \mathbf{obj}(\phi, q)$. The next lemma gives a sufficient condition for $\phi \models q$ iff $\mathbf{obj}(\phi) \models \mathbf{obj}(\phi, q)$, where we write $\text{Var}(\tau)$ for the set of all free

variables in a formula τ , $M(\text{Var}(\tau), \mathbf{D})$ for the set of all mappings from $\text{Var}(\tau)$ to \mathbf{D} , and $v(\mathcal{I}, \mathbf{I}_V)$ for an interpretation that coincides with \mathbf{I}_V on all variables it interprets, and with \mathcal{I} on everything else.

We say that a formula has a *positive* occurrence if it is in a fact or a rule conclusion, and a *negative* occurrence if in a query or a rule premise. A semantic structure \mathcal{I} is called a *counter-model* for $\phi \models q$ if $\mathcal{I} \models \phi$ and $\mathcal{I} \not\models q$.

Lemma 1. *Given a KB ϕ , a query q , and a transformation obj , $\phi \models q$ iff $\text{obj}(\phi) \models \text{obj}(\phi, q)$ holds if obj has the following properties:*

- 1) *For every counter-model \mathcal{I} for $\phi \models q$, there exists \mathcal{I}' s.t. for every KB / query atom ω , $\mathbf{I}_V \in M(\text{Var}(\omega), \mathcal{I}.\mathbf{D})$,

 - 1.a) *if ω is positive and $v(\mathcal{I}, \mathbf{I}_V) \models \omega$, then $v(\mathcal{I}', \mathbf{I}_V) \models \text{obj}(\phi, \omega)$;*
 - 1.b) *if ω is negative and $v(\mathcal{I}', \mathbf{I}_V) \models \text{obj}(\phi, \omega)$, then $v(\mathcal{I}, \mathbf{I}_V) \models \omega$.**
- 2) *For every counter-model \mathcal{I}' for $\text{obj}(\phi) \models \text{obj}(\phi, q)$, there exists \mathcal{I} s.t. for every KB / query atom ω , $\mathbf{I}_V \in M(\text{Var}(\omega), \mathcal{I}'.\mathbf{D})$,

 - 2.a) *if ω is positive and $v(\mathcal{I}', \mathbf{I}_V) \models \text{obj}(\phi, \omega)$, then $v(\mathcal{I}, \mathbf{I}_V) \models \omega$,*
 - 2.b) *if ω is negative and $v(\mathcal{I}, \mathbf{I}_V) \models \omega$, then $v(\mathcal{I}', \mathbf{I}_V) \models \text{obj}(\phi, \omega)$.**

Proof. (Sketch) Proving $\phi \models q$ iff $\text{obj}(\phi) \models \text{obj}(\phi, q)$ is equivalent to proving $\phi \not\models q$ iff $\text{obj}(\phi) \not\models \text{obj}(\phi, q)$. We first prove the ‘if’ part. If $\text{obj}(\phi) \not\models \text{obj}(\phi, q)$, then there exists a counter-model \mathcal{I}' for $\text{obj}(\phi) \models \text{obj}(\phi, q)$. Using condition 2.a), we can prove by induction that for every positive formula τ and $\mathbf{I}_V \in M(\text{Var}(\tau), \mathcal{I}.\mathbf{D})$, if $TVal_{v(\mathcal{I}', \mathbf{I}_V)}(\text{obj}(\phi, \tau)) = \mathbf{t}$ then $TVal_{v(\mathcal{I}, \mathbf{I}_V)}(\tau) = \mathbf{t}$. Using condition 2.b), we can also prove by induction that for every negative formula τ , if $TVal_{v(\mathcal{I}, \mathbf{I}_V)}(\tau) = \mathbf{t}$ then $TVal_{v(\mathcal{I}', \mathbf{I}_V)}(\text{obj}(\phi, \tau)) = \mathbf{t}$.

Then we can prove that for each top-level formula τ in ϕ , $TVal_{\mathcal{I}}(\text{obj}(\phi, \tau)) = \mathbf{t}$.

- If τ is a ground fact, τ is a positive formula so $TVal_{\mathcal{I}}(\text{obj}(\phi, \tau)) = \mathbf{t}$.
- If τ is a rule of the form $\text{For all } ?X_1 \dots ?X_n (\tau_1 :- \tau_2)$, $TVal_{\mathcal{I}'}(\tau) = \mathbf{t}$ means for every $\mathbf{I}_V^* \in M(\{?X_1, \dots, ?X_n\}, \mathcal{I}.\mathbf{D})$, $TVal_{v(\mathcal{I}', \mathbf{I}_V^*)}(\tau_1) = \mathbf{t}$ or $TVal_{v(\mathcal{I}', \mathbf{I}_V^*)}(\tau_2) = \mathbf{f}$ holds. Since τ_1 is positive and τ_2 is negative, either $TVal_{v(\mathcal{I}, \mathbf{I}_V^*)}(\tau_1) = \mathbf{t}$ or $TVal_{v(\mathcal{I}, \mathbf{I}_V^*)}(\tau_2) = \mathbf{f}$ holds for every \mathbf{I}_V^* . So $TVal_{\mathcal{I}}(\tau) = \mathbf{t}$.
- If τ is a non-ground fact, it can be seen as a rule with an empty premise, so that the above proof for general rules still apply.

Hence $\mathcal{I} \models \phi$. Also since the query q is negative, hence $\mathcal{I} \not\models q$. Thus $\phi \not\models q$.

The ‘only if’ part can be proved using a similar approach based on the existence of a counter-model \mathcal{I}' for $\text{obj}(\phi) \models \text{obj}(\phi, q)$. \square

Corollary 1. *Given a KB ϕ , a query q , and a transformation obj , $\phi \models q$ iff $\text{obj}(\phi) \models \text{obj}(\phi, q)$ holds if for every KB / query atom ω and semantic structure \mathcal{I}^* , $\mathcal{I}^* \models \omega$ iff $\mathcal{I}^* \models \text{obj}(\phi, \omega)$ holds.*

Proof. We will show that obj has the properties 1) and 2) in Lemma 1. For property 1), if \mathcal{I} is a counter-model for $\phi \models q$, then we can choose $\mathcal{I}' = \mathcal{I}$. For every $\mathbf{I}_V \in M(\text{Var}(\omega), \mathcal{I}.\mathbf{D})$, according to the assumption where \mathcal{I}^* becomes $v(\mathcal{I}, \mathbf{I}_V)$, we have $v(\mathcal{I}, \mathbf{I}_V) \models \omega$ iff $v(\mathcal{I}, \mathbf{I}_V) \models \text{obj}(\phi, \omega)$. Hence 1.a) and 1.b) are satisfied and obj has property 1).

That obj has property 2) can be proved similarly by choosing $\mathcal{I} = \mathcal{I}'$. \square

Lemma 2 gives a sufficient condition for $\text{obj}(\phi) \models \text{obj}(\phi, q)$ iff $\text{obj}(\phi) \models_{-o} \text{obj}(\phi, q)$.

Lemma 2. *Given a KB ϕ' and a query q' , $\phi' \models q'$ iff $\phi' \models_{-o} q'$ holds if for each counter-model \mathcal{I}'' for $\phi' \models_{-o} q'$ there exists a counter-model \mathcal{I}' for $\phi' \models q'$.*

Proof. We first prove the ‘if’ part. For every \mathcal{I}' s.t. $\mathcal{I}' \models \phi'$, $\mathcal{I}' \models_{-o} \phi'$ since \models more restricted than \models_{-o} . Also since $\phi' \models_{-o} q'$, $\mathcal{I}' \models_{-o} q'$. Because $\mathcal{I}' \models \phi'$, \mathcal{I}' guarantees the objectification restriction. Hence $\mathcal{I}' \models q'$ always holds and $\phi' \models q'$ is proved.

Next we prove the ‘only if’ part, which is equivalent to $\phi' \not\models q'$ if $\phi' \not\models_{-o} q'$. If $\phi' \not\models_{-o} q'$, there exists a counter-model \mathcal{I}'' for $\phi' \models_{-o} q'$. According to the assumption of the lemma, there exists a counter-model \mathcal{I}' for $\phi' \models q'$, hence $\phi' \not\models q'$ is proved. \square

The next subsections will discuss different objectification transformations.

3.2 Static Objectification Transformations

The static objectification $\text{obj}_s(\alpha)$ of a KB/query α is obtained by replacing each oidless atom σ with its objectified form $\text{obj}_s(\sigma)$ having a generated OID. The generation can adopt either an undifferentiated method $\text{obj}_{s=}$, which uniformly transforms σ everywhere, or a differentiated method $\text{obj}_{s\neq}$ [6], which transforms σ differently based on its occurrence.

Definition 4. (*Undifferentiated static objectification*) *The undifferentiated static objectification $\text{obj}_{s=}(\omega)$ of atom is (a) ω if ω is oidful; (b) **Exists ?i** ($?i\#f(\dots)$) if ω is oidless, where ?i is a fresh variable in the clause chosen from ?1, ?2, ... $\text{obj}_{s=}$ is extended for other formulas according to Definition 3.*

In Example 1, $\text{obj}_{s=}$ of the KB and the queries replaces all oidless `_transfer`, `_work`, and `_acquire` atoms with their existential forms.

The following theorem shows the correctness of $\text{obj}_{s=}$.

Theorem 1. *$\text{obj}_{s=}$ is an objectification transformation for any KB and query.*

Proof. We first show that for every KB / query atom ω and semantic structure \mathcal{I}^* , $\mathcal{I}^* \models \omega$ iff $\mathcal{I}^* \models \text{obj}_{s=}(\omega)$ holds. If ω is oidful, then $\text{obj}_{s=}(\omega) = \omega$, and the statement holds trivially. Otherwise ω is oidless. If $\mathcal{I}^* \models \omega$, then \mathcal{I}^* guarantees the objectification restriction. So $TVal_{\mathcal{I}^*}(\text{obj}_{s=}(\omega)) = TVal_{\mathcal{I}^*}(\omega) = \mathbf{t}$, and $\mathcal{I}^* \models \text{obj}_{s=}(\omega)$ holds. Similarly, if $\mathcal{I}^* \not\models \text{obj}_{s=}(\omega)$, $\mathcal{I}^* \not\models \omega$ also holds. Hence $\mathcal{I}^* \models \omega$ iff $\mathcal{I}^* \models \text{obj}_{s=}(\omega)$ holds also for oidless atoms. Thus the condition of Corollary 1 is fulfilled and $\phi \models q$ iff $\text{obj}_{s=}(\phi) \models \text{obj}_{s=}(q)$.

Next we will show that for each $\text{obj}_{s=}(\phi)$ and $\text{obj}_{s=}(q)$, the condition of Lemma 2 is fulfilled. If there exists a counter-model \mathcal{I}'' for $\phi \models q$, then it can be made into a model \mathcal{I}' that guarantees the objectification restriction, by redefining $TVal_{\mathcal{I}'}(\omega)$ for every atom ω that is oidless to be the same as $TVal_{\mathcal{I}''}(\text{obj}_{s=}(\omega))$. Since the change only affect oidless atoms, which neither exist in $\text{obj}_{s=}(\phi)$ nor

in $\text{obj}_{s=}(q)$, $TVal_{\mathcal{T}'}(\phi) = TVal_{\mathcal{T}''}(\phi) = \mathbf{t}$ and $TVal_{\mathcal{T}'}(q) = TVal_{\mathcal{T}''}(q) = \mathbf{f}$. Hence \mathcal{T}' is a counter-model for $\phi \models q$ and the condition of Lemma 2 is fulfilled. Thus $\text{obj}_{s=}(\phi) \models \text{obj}_{s=}(q)$ iff $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$.

So $\phi \models q$ iff $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$ and the theorem is proved. \square

Definition 5. (*Differentiated static objectification*) For an atom ω , the differentiated static objectification $\text{obj}_{s\neq}(\omega)$ is defined as ω if ω is oidful, or as follows if ω is oidless:

Case 1: If ω is a ground fact, $\text{obj}_{s\neq}(\omega) = _i \# \mathbf{f}(\dots)$, where $_i$ is a fresh local constant symbol chosen from $_1, _2, \dots$, which neither occurs elsewhere in the KB nor is used for the objectification of other atoms.

Case 2: If ω is a non-ground fact, a rule conclusion atom, or a query atom, $\text{obj}_{s\neq}(\omega) = \text{Exists } ?j \ (?j \# \mathbf{f}(\dots))$.

Case 3: If ω is a rule premise atom, $\text{obj}_{s\neq}(\omega) = ?j \# \mathbf{f}(\dots)$, where $?j$ is a fresh variable scoped universally by the enclosing rule.

$\text{obj}_{s\neq}$ is extended for other formulas according to Definition 3.

In Example 1, differentiated static objectification generates a fresh OID constant for the `_work` ground fact according to Case 1, replaces the oidless `_work` and `_transfer` atoms in the rule conclusion and in the queries with existentials according to Case 2, and generates fresh universal OID variables for the `_work` and `_acquire` atoms in the rule premise according to Case 3.

It is easy to prove the correctness of $\text{obj}_{s\neq}$ by showing that it leads to the same set of entailed queries as $\text{obj}_{s=}$.

Lemma 3. Given a KB ϕ and a query q , if q does not use a generated OID constant symbol $_1, _2, \dots$, then $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$ iff $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s\neq}(q)$.

Proof. According to their definitions, both methods transform oidless query atoms into the same existential form, hence $\text{obj}_{s=}(q) = \text{obj}_{s\neq}(q)$ for any query q . Thus $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s=}(q)$ iff $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s\neq}(q)$.

For KB atoms, the two methods are identical in Case 2 and differ in Case 1 and 3. The transformation of a rule premise atom in Case 3 using a universal OID variable on the top-level is equivalent to using an embedded existentially quantified formula with an existential OID variable in the rule premise. For a ground fact ω handled by Case 1, $\text{obj}_{s\neq}(\omega)$ can be seen as a Skolemized version of $\text{obj}_{s=}(\omega)$ using Skolem constants $_1, _2, \dots$, hence they entail the same set of queries as long as these queries do not use the Skolem constants, thus avoiding a clash of constant symbols. Hence we have $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$ iff $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s=}(q)$. So $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$ iff $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s=}(q)$, and the lemma holds. \square

Theorem 2. $\text{obj}_{s=}$ is an objectification transformation with respect to a set of KBs Φ and to a set of queries Q that do not use constants $_1, _2, \dots$

Proof. For any $\phi \in \Phi$ and $q \in Q$, by Theorem 1 we have $\phi \models q$ iff $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$. By Lemma 3, $\text{obj}_{s=}(\phi) \models_{-o} \text{obj}_{s=}(q)$ iff $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s\neq}(q)$. Hence $\phi \models q$ iff $\text{obj}_{s\neq}(\phi) \models_{-o} \text{obj}_{s\neq}(q)$. So the statement holds by Definition 2. \square

3.3 Static/Dynamic Objectification Transformation

For KBs in which most or all of the predicates are Prolog-like relations, it is often not necessary to generate OIDs for their oidless atoms explicitly. In this subsection, a novel static/dynamic objectification approach is introduced to keep unchanged as many of the KB's oidless atoms as possible, instead constructing virtual OIDs at query time when bindings for OID variables are being queried.

In order to apply static/dynamic objectification to a KB ϕ and its queries, the set of KB predicates $\text{PredKB}(\phi)$ will be partitioned into two disjoint subsets. $\text{PredKB}(\phi)$ is defined as $\{\text{Pred}(\lambda) \mid \lambda \text{ is an atom in } \phi\}$, where $\text{Pred}(\lambda)$ denotes the predicate symbol of λ . The partitioning of $\text{PredKB}(\phi)$ is defined next.

Definition 6. (*Non-relational and relational predicates*) Given a KB ϕ , a predicate $\mathbf{f} \in \text{PredKB}(\phi)$ is **non-relational** in ϕ if \mathbf{f} occurs at least once in a multi-tuple, oidful, or slotted atom of ϕ , or in a subclass formula of ϕ . Conversely, \mathbf{f} is **relational** in ϕ if it has no such occurrence. The sets of non-relational and relational predicates of ϕ are written as $\text{PredKB}_{NR}(\phi)$ and $\text{PredKB}_R(\phi)$, respectively.

For atoms using a relational predicate in $\text{PredKB}_R(\phi)$, their OIDs can be virtualized by dynamic objectification:

Definition 7. (*Dynamic objectification*) The dynamic objectification $\text{obj}_d(\phi, \omega)$ of an atom ω with respect to a KB ϕ , where $\text{Pred}(\omega) \in \text{PredKB}_R(\phi)$, is defined as ω if ω is a KB atom in ϕ , or as the following rewriting if ω is a query atom:

- Case 1: If ω is a relationship, $\text{obj}_d(\phi, \omega) = \omega$.
- Case 2: If ω has a non-variable (e.g., constant or expression) OID or a slot, $\text{obj}_d(\phi, \omega) = \text{Or}()$, where $\text{Or}()$ is an encoding of explicit falsity.
- Case 3: If ω has an OID variable and $m > 0$ tuples, being of the form $?0\#f([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}] \dots [\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}])$, equivalent (via ‘class-reproducing’ tupribution) to a conjunction separately applying $?0\#f$ to all tuples,

$$\text{And}(?0\#f([\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}]) \dots ?0\#f([\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}])),$$

abbreviated to

$$\text{And}(?0\#f(\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}) \dots ?0\#f(\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m})),$$

then $\text{obj}_d(\phi, \omega)$ is a relational conjunction querying the oidless versions of these applications while using explicit equalities between the OID variable $?0$ and an OID-constructor function $_oidcons$ applied to the predicate and the elements of each tuple (where the $?0$ equalities also enforce tuple unification, so that the m single-tuple \mathbf{f} relationships can be satisfied by a single KB clause, thus realizing special ‘virtual multi-tuple’ psoa atoms as queries):

$$\begin{aligned} &\text{And}(f(\mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}) ?0 = _oidcons(f \mathbf{t}_{1,1} \dots \mathbf{t}_{1,n_1}) \\ &\dots \\ &f(\mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m}) ?0 = _oidcons(f \mathbf{t}_{m,1} \dots \mathbf{t}_{m,n_m})) \end{aligned}$$

In the special case of $m = 1$, the query atom $?0\#f(t_1 \dots t_n)$ becomes

$$\text{And}(f(t_1 \dots t_n) ?0 = _oidcons(f t_1 \dots t_n))$$

The function $_oidcons$ is employed to universally construct OIDs for all relationships so that it needs to take the predicate symbol as the first argument to distinguish between relationships with different predicates.

Case 4: If ω is a membership of the form $?0\#f$, $\text{obj}_d(\phi, \omega)$ is a disjunction of k formulas $\text{obj}_d(?0\#f(?X_1 \dots ?X_{n_k}))$, where n_1, \dots, n_k are the k different arities of f in the KB:

$$\text{Or}(\text{obj}_d(?0\#f(?X_1 \dots ?X_{n_1})) \dots \text{obj}_d(?0\#f(?X_1 \dots ?X_{n_k})))$$

Case 5: If ω of the form $f(\dots)$ has no OID but m tuples, $m > 1$, $\text{obj}_d(\phi, \omega) = \text{Exists } ?0 \text{ obj}_d(?0\#f(\dots))$, where $?0$ is a fresh variable in the query.

Definition 8. (Static/dynamic objectification) Let obj_s be a transformation chosen from $\{\text{obj}_{s \neq}, \text{obj}_{s =}\}$ and ϕ be a KB. The static/dynamic objectification $\text{obj}_{s+d}(\phi, \omega)$ of an atom is $\text{obj}_s(\phi, \omega)$ if $\text{Pred}(\omega) \in \text{PredKB}_{NR}(\phi)$, or $\text{obj}_d(\phi, \omega)$ if $\text{Pred}(\omega) \in \text{PredKB}_R(\phi)$. obj_{s+d} is extended for the KB ϕ and other formulas, using Definition 3, if the following conditions are satisfied:

- (i) For every KB clause, universal variables occurring in its conclusion must also occur in its premise (e.g., prohibiting non-ground facts).
- (ii) There does not exist a predicate variable or an untyped (*Top*-typed) atom with positional arguments in the premise of a rule or in the top-level query.

In Example 1, conditions (i) and (ii) are satisfied so that static/dynamic objectification can be applied. In the KB, the predicate $_work$ is relational while the others are non-relational. Thus, all oidless $_work$ atoms are kept unchanged while for the other oidless atoms OIDs are introduced via one of the two static methods. The query $?0\#_work(?P ?C ?J)$ is rewritten into the conjunction $\text{And}(_work(?P ?C ?J) ?0 = _oidcons(_work ?P ?C ?J))$ according to Case 3 of Definition 7.

In the following we will show the correctness of $\text{obj}_{s=+d}$ and $\text{obj}_{s \neq +d}$, which correspond to the two obj_{s+d} versions using $\text{obj}_{s=}$ and $\text{obj}_{s \neq}$ for the static part.

Lemma 4. Let ϕ be a KB and q be a query without $_oidcons$ and let ϕ and q comply to the above conditions (i) and (ii). Then $\text{obj}_{s=+d}$ has the properties (1) and (2) of Lemma 1 with respect to ϕ and q .

Proof. (Sketch) If \mathcal{I} is a counter-model for $\phi \models q$, we can define \mathcal{I}' to be modified from \mathcal{I} by adding interpretation of $_oidcons$ so that for every relationship ω where $\text{Pred}(\omega) \in \text{PredKB}_R(\phi)$, being of the form $f(t_1 \dots t_n)$, $TVal_{\mathcal{I}'}(_oidcons(f t_1 \dots t_n)\#f(t_1 \dots t_n)) = \mathbf{t}$. This is done by making $\mathbf{I}(_oidcons(f t_1 \dots t_n))$ the domain element that represents the OID of ω . We can prove that \mathcal{I}' fulfills the conditions 1.a) and 1.b) for every KB / query atom ω and $\mathbf{I}_V \in \mathbf{M}(\text{Var}(\omega), \mathcal{I}.D)$ performing a case-by-case analysis for ω . We can also prove that $\text{obj}_{s=+d}$ satisfies conditions 2) by choosing $\mathcal{I} = \mathcal{I}'$ and doing a case-by-case analysis. \square

Lemma 5. *Let ϕ be a KB and q be a query without `_oidcons`, let ϕ and q comply to the above conditions (i) and (ii), and let $\phi' = \text{obj}_{s=+d}(\phi)$ and $q' = \text{obj}_{s=+d}(\phi, q)$. Then ϕ' and q' fulfill the conditions of Lemma 2.*

Proof. (Sketch) If \mathcal{I}'' is a counter-model for $\phi' \models_{-o} q'$, then $\mathcal{I}'' \models_{-o} \phi'$ and $\mathcal{I}'' \not\models_{-o} q'$. We modify \mathcal{I}'' into a semantic structure \mathcal{I}' that guarantees the objectification restriction using the following redefinitions:

- a) For every oidful atom ω' in ϕ' and q' , we redefine $TVal_{\mathcal{I}'}(\omega)$ for its oidless form ω to be the same as $TVal_{\mathcal{I}''}(\omega')$.
- b) For every oidless atom ω' in ϕ' and q' of the form $\mathbf{f}(\mathbf{t}_1 \dots \mathbf{t}_n)$, let $\mathbf{o}(\omega')$ be the virtual OID `_oidcons(f t1 ... tn)`; we redefine $TVal_{\mathcal{I}'}(\mathbf{o}(\omega')\#\mathbf{f}(\mathbf{t}_1 \dots \mathbf{t}_n))$, $TVal_{\mathcal{I}'}(\mathbf{o}(\omega')\#\mathbf{f})$, and $TVal_{\mathcal{I}'}(\mathbf{o}(\omega')\#\text{Top}(\mathbf{t}_1 \dots \mathbf{t}_n))$ to be the same as $TVal_{\mathcal{I}''}(\omega')$. All other atoms that have a component $\mathbf{o}(\omega')$ are required to evaluate to false.

If $\text{Pred}(\omega') \in \text{PredKB}_{NR}(\phi')$, definition a) applies, and the redefinition would not affect the truth evaluation for formulas in ϕ' and q' since ω does not occur in ϕ' .

If $\text{Pred}(\omega') \in \text{PredKB}_R(\phi')$, definition b) applies, and the objectification restriction is guaranteed. We will show that the redefinition would not affect the truth evaluation for top-level formulas in ϕ' and q' . Since $\mathbf{f} \in \text{PredKB}_R(\phi')$, it does not occur in an oidful or multi-tuple atom. So \mathcal{I}' preserves the truth evaluation of all atoms with the predicate \mathbf{f} . We discuss different top-level formulas τ in ϕ' :

- For τ being a ground fact, if $\text{Pred}(\tau) \in \text{PredKB}_R(\phi')$ then $TVal_{\mathcal{I}'}(\tau) = TVal_{\mathcal{I}''}(\tau) = \mathbf{t}$. Otherwise, its components cannot be interpreted to any $\mathbf{o}(\omega')$ so $TVal_{\mathcal{I}'}(\tau) = TVal_{\mathcal{I}''}(\tau) = \mathbf{t}$ still holds.
- τ being a non-ground fact is disallowed by condition (i).
- For τ being a rule `forall ?X1 ... ?Xn ($\tau_1 :- \tau_2$)`, for every $\mathbf{I}_V \in \mathbf{M}(\{\text{?X}_1, \dots, \text{?X}_n\}, \mathcal{I}''.\mathbf{D})$, if for some i , $\mathbf{I}_V(\text{?X}_i)$ interprets to any $\mathbf{o}(\omega')$, then it can be shown that $TVal_{\mathbf{v}(\mathcal{I}', \mathbf{I}_V)}(\tau_2) = \mathbf{f}$ always holds based on Conditions (i) and (ii). So $TVal_{\mathbf{v}(\mathcal{I}', \mathbf{I}_V)}(\tau_1 :- \tau_2) = \mathbf{t}$. Otherwise, $TVal_{\mathbf{v}(\mathcal{I}', \mathbf{I}_V)}(\tau_1 :- \tau_2) = TVal_{\mathbf{v}(\mathcal{I}'', \mathbf{I}_V)}(\tau_1 :- \tau_2) = \mathbf{t}$. Hence $TVal_{\mathcal{I}'}(\tau) = \mathbf{t}$.

\mathcal{I}' can also be verified to conform to the subclass and slotribution/tupribution restrictions. So \mathcal{I}' is a counter-model for $\phi' \models q'$, and the lemma holds. \square

Theorem 3. *$\text{obj}_{s=+d}$ is an objectification transformation with respect to a set of KBs Φ and a set of queries Q that do not use `_oidcons` and satisfy conditions (i) and (ii).*

Proof. By Lemmas 4, $\phi \models q$ iff $\text{obj}_{s=+d}(\phi) \models \text{obj}_{s=+d}(\phi, q)$. By Lemmas 5 and 5, $\text{obj}_{s=+d}(\phi) \models \text{obj}_{s=+d}(\phi, q)$ iff $\text{obj}_{s=+d}(\phi) \models_{-o} \text{obj}_{s=+d}(\phi, q)$. Hence $\phi \models q$ iff $\text{obj}_{s=+d}(\phi) \models_{-o} \text{obj}_{s=+d}(\phi, q)$ and the theorem is proved. \square

Theorem 4. *$\text{obj}_{s \neq +d}$ is an objectification transformation with respect to KBs Φ and queries Q that do not use constants `_oidcons`, `_1`, `_2`, ... and satisfy conditions (i) and (ii).*

Proof. (Sketch) This can be proved similarly to Theorem 2 by first proving a corresponding version of Lemma 3 for $\text{obj}_{s \neq +d}$ and $\text{obj}_{s=+d}$. \square

4 Maximal Unnesting

In this section we first discuss the syntax and semantics of embedded psOA terms and then formally define the unnesting transformation of nested atomic formulas. Unnesting is maximal in the sense that it can extract atoms not only from other atoms but also from expressions, which may themselves be embedded at any level.

4.1 Syntax and Semantics of Embedded PsOA Terms

In PSOA RuleML, embedded psOA terms can be semantically classified into expressions or atoms, and syntactically classified into oidful or oidless terms. We will discuss different combinations of the two dimensions in the following.

Expressions A psOA expression is a psOA term interpreted as a function application. Like in classical logic and logic programming languages, we require a (psOA) expression to be oidless, having the form $\mathbf{f}(\dots)$, with \mathbf{f} acting as a function, hence leading to an arbitrary value. The expression cannot be given an OID and have the form $\mathbf{o}\#\mathbf{f}(\dots)$, because this would make \mathbf{f} act the class of \mathbf{o} , hence lead to a truth value.

The semantics for expressions uses the \mathbf{I} and \mathbf{I}_{psOA} mappings defined in Section 3.1. However, because these expressions must be oidless, \mathbf{I}_{psOA} interprets $\mathbf{I}(\mathbf{f})$ into a semantic function that takes the empty set for oidless psOA terms, besides a bag of tuples and a bag of slots, and returns an arbitrary domain element. The lack of an OID for an expression also prevents its slottribution/tuptribution, thus avoiding that its arbitrary value gets ‘distributed over’ the resulting conjunction of single-slot/tuple expressions or ‘absorbed by’ its truth value.⁴

Atoms A psOA atom is a psOA term interpreted as a predicate application. Embedded atoms have been widely used in object-centered languages such as RDF, N3 [8], and Flora-2/F-logic [9] as a shorthand notation. An atomic formula containing an embedded atom can be unnested into a conjunction of trimmed formulas, which will be illustrated in Section 4.2.

In RDF and N3, embedded atoms are oidless and are ‘objectified-while-embedded’ using blank nodes, while in Flora-2/F-logic they are oidful. Because in PSOA/PS the set **Const** of constants contains both functions and predicates (cf. Section 2), an oidless embedded atom cannot be distinguished from an expression through the alphabet, hence we use the syntactic distinction between an embedded oidless term for an expression and an embedded oidful term for an atom. The distinction applies only to embedded psOA terms but not to top-level terms, where both oidful and oidless forms are interpreted as atoms. To indicate an RDF/N3-like embedded blank node, the atom can use an explicit OID,

⁴ Relfun’s *valued conjunctions* of functional-logic expressions only retain the value of the right-most, ‘&’-prefixed conjunct: <http://www.relfun.org>.

thus: (1) $_#f(\dots)$ in ground facts, where ‘ $_$ ’ is the fresh-constant generator; (2) $\text{Exists } ?0(\dots ?0\#f(\dots) \dots)$ in non-ground facts and rule conclusions; and (3) $?#f(\dots)$ in queries / rule premises, where ‘?’ is the anonymous variable.

4.2 Unnesting Transformation for Embedded Psoa Atoms

In this section we will define the unnesting transformation $\text{Unnest}(\alpha)$ for a given atomic formula α , which is extended from the definition in [10]. Before performing $\text{Unnest}(\alpha)$, anonymous OID constants and variables in α need to be eliminated, because they cannot be used as co-references for the same constant/variable in two separate formulas. The elimination can be done by replacing ‘ $_$ ’ with a fresh constant and ‘?’ with a fresh variable.

In the following we give the definition of $\text{Unnest}(\alpha)$ based on the recursive Atoms . Here, $\text{Oid}(t)$ denotes the OID of an oidful term t . Also, $\text{Parts}(t)$ denotes the set of top-level components of an atomic formula or a term t , including, optionally, $\text{Oid}(t)$, its positional arguments, slot names, slot fillers, as well as its predicate/function.

$$\begin{aligned} \text{Unnest}(\alpha) &::= \text{And}(\sigma_1 \dots \sigma_n) \quad \text{s.t. } \{\sigma_1, \dots, \sigma_n\} = \bigcup_{t \in \text{Parts}(\alpha)} \text{Atoms}(t) \bigcup \{\text{Trim}(\alpha)\} \\ \text{Atoms}(t) &::= \begin{cases} \emptyset & t \text{ is a simple term} \\ \bigcup_{s \in \text{Parts}(t)} \text{Atoms}(s) & t \text{ is oidless (expression)} \\ \bigcup_{s \in \text{Parts}(t)} \text{Atoms}(s) \bigcup \{\text{Trim}(t)\} & t \text{ is oidful (atom)} \end{cases} \\ \text{Trim}(t) &::= \text{Term/Formula obtained by replacing every } s \in \text{Parts}(t) \text{ in } t \text{ with } \text{Retain}(s) \\ \text{Retain}(t) &::= \begin{cases} t & t \text{ is a simple term} \\ \text{Trim}(t) & t \text{ is oidless (expression)} \\ \text{Retain}(\text{Oid}(t)) & t \text{ is oidful (atom)} \end{cases} \end{aligned}$$

$\text{Unnest}(\alpha)$ is a conjunction of formulas σ_i without embedded atoms. Each σ_i is a trimmed version of the top-level formula α or of some embedded psOA atom. The set $\text{Atoms}(t)$ contains each σ_i trimmed from an atom embedded in t or t itself. It is constructed by recursively traversing through each component $s \in \text{Parts}(t)$, collecting $\text{Atoms}(s)$ into $\text{Atoms}(t)$, and then adding $\text{Trim}(t)$ to $\text{Atoms}(t)$ if t is oidful, which indicates that t is an atom. The transformation $\text{Trim}(t)$ splits off all embedded atoms from t and leaves behind its ‘ultimate’ OID for each of them. It is constructed by replacing each $s \in \text{Parts}(t)$ with $\text{Retain}(s)$, which defines the left-behind term for each embedded term s .

In the following we use an example to explain the unnesting transformation. Let the input formula α be $\text{o1}\#c(\text{p} \rightarrow \text{f}(\text{o2}\#c\#d))$. Note that ‘#’ is left-associative, hence the embedded atom $\text{o2}\#c\#d$ is interpreted to have the OID $\text{o2}\#c$ and the class d . The conjuncts of $\text{Unnest}(\alpha)$ are constructed as follows:

$$\begin{aligned} \{\sigma_1, \dots, \sigma_n\} &= \bigcup_{t \in \text{Parts}(\alpha)} \text{Atoms}(t) \bigcup \{\text{Trim}(\alpha)\} \\ &= (\text{Atoms}(\text{o1}) \cup \text{Atoms}(c) \cup \text{Atoms}(\text{p}) \cup \text{Atoms}(\text{f}(\text{o2}\#c\#d))) \bigcup \{\text{Trim}(\alpha)\} \\ &= \text{Atoms}(\text{f}(\text{o2}\#c\#d)) \bigcup \{\text{Trim}(\alpha)\} \\ &= (\text{Atoms}(\text{f}) \cup \text{Atoms}(\text{o2}\#c\#d)) \bigcup \{\text{Trim}(\alpha)\} \end{aligned}$$

$$\begin{aligned}
&= \text{Atoms}(\text{o2\#c\#d}) \bigcup \{\text{Trim}(\alpha)\} \\
&= (\text{Atoms}(\text{o2\#c}) \cup \text{Atoms}(\text{d}) \cup \{\text{Trim}(\text{o2\#c\#d})\}) \bigcup \{\text{Trim}(\alpha)\} \\
&= \text{Atoms}(\text{o2\#c}) \bigcup \{\text{Trim}(\text{o2\#c\#d})\} \bigcup \{\text{Trim}(\alpha)\} \\
&= (\text{Atoms}(\text{o2}) \cup \text{Atoms}(\text{c}) \cup \{\text{Trim}(\text{o2\#c})\}) \bigcup \{\text{Trim}(\text{o2\#c\#d})\} \bigcup \{\text{Trim}(\alpha)\} \\
&= \{\text{Trim}(\text{o2\#c}), \text{Trim}(\text{o2\#c\#d}), \text{Trim}(\alpha)\}
\end{aligned}$$

The Trim transformations work as follows, using the recursive Retain transformation:

$$\begin{aligned}
\text{Trim}(\text{o2\#c}) &= \text{Retain}(\text{o2})\#\text{Retain}(\text{c}) = \text{o2\#c} \\
\text{Trim}(\text{o2\#c\#d}) &= \text{Retain}(\text{o2\#c})\#\text{Retain}(\text{d}) \\
&= \text{Retain}(\text{Oid}(\text{o2\#c}))\#\text{d} = \text{Retain}(\text{o2})\#\text{d} = \text{o2\#d} \\
\text{Trim}(\alpha) &= \text{Trim}(\text{o1\#c}(\text{p}\rightarrow\text{f}(\text{o2\#c\#d}))) \\
&= \text{Retain}(\text{o1})\#\text{Retain}(\text{c})(\text{Retain}(\text{p})\rightarrow\text{Retain}(\text{f}(\text{o2\#c\#d}))) \\
&= \text{o1\#c}(\text{p}\rightarrow\text{Retain}(\text{f})(\text{Retain}(\text{o2\#c\#d}))) \\
&= \text{o1\#c}(\text{p}\rightarrow\text{f}(\text{Retain}(\text{Oid}(\text{o2\#c\#d})))) \\
&= \text{o1\#c}(\text{p}\rightarrow\text{f}(\text{Retain}(\text{o2\#c}))) \\
&= \text{o1\#d}(\text{p}\rightarrow\text{f}(\text{o2}))
\end{aligned}$$

Hence, the unnesting $\text{Unnest}(\alpha)$ results in $\text{And}(\text{o2\#c} \ \text{o2\#d} \ \text{o1\#c}(\text{p}\rightarrow\text{f}(\text{o2})))$.

The unnesting transformation has been implemented in the latest release of PSOATransRun 1.1 using a separate ANTLR tree walker.

5 Conclusions

This paper discusses advanced objectification and unnesting transformations for the PSOA RuleML language as well as a novel model-theoretic semantics.

The refined semantics is introduced to allow a direct interpretation of oidless psoa terms. It includes the objectification restriction to establish the equivalence between an oidless atom and its existentially objectified form. Based on the new semantics, a systematics of three objectification transformations is defined, whose correctness is proved. A novel static/dynamic objectification approach for oidless atoms is introduced, which is minimal in that it generates as few explicit OIDs as possible, instead constructing virtual OIDs as query variable bindings. This approach provides better efficiency for the PSOATransRun implementation by allowing direct use of the underlying Prolog engine.

The unnesting transformation is formalized to decompose nested atomic formulas into equivalent conjunctions before applying the model-theoretic semantics. Unnesting is maximal in that it can recursively extract oidful atoms – leaving behind their OIDs – not only from other atoms but also from expressions, which may themselves be embedded at any level. Since embedded oidless terms are interpreted as expressions (usable, e.g., as ‘passive’ data constructors), rather than as atoms to be ‘objectified-while-embedded’, for embedded

atoms OIDs need to be explicitly provided to enable unnesting. The unnesting transformation has been implemented in PSOATransRun 1.1.

Future work includes exploring further optimizations for objectification and complementing the unnesting transformation with a flattening transformation for extracting ‘active’ expressions from both atoms and expressions. For this, PSOATransRun’s flattening of expressions calling built-in functions can be easily transferred to expressions with equality-defined functions.

References

1. Kifer, M., Lausen, G., Wu, J.: Logical foundations of object-oriented and frame-based languages. *Journal of the ACM* **42**(4) (July 1995) 741–843
2. Yang, G., Kifer, M.: Reasoning about Anonymous Resources and Meta Statements on the Semantic Web. In Spaccapietra, S., March, S.T., Aberer, K., eds.: *J. Data Semantics I*. Volume 2800 of *Lecture Notes in Computer Science.*, Springer (2003) 69–97
3. Boley, H., Kifer, M.: *RIF Basic Logic Dialect (Second Edition)* (February 2013) W3C Recommendation, <http://www.w3.org/TR/rif-bld>.
4. Boley, H.: A RIF-Style Semantics for RuleML-Integrated Positional-Slotted, Object-Applicative Rules. In: *Proc. 5th International Symposium on Rules: Research Based and Industry Focused (RuleML-2011 Europe)*, Barcelona, Spain. *Lecture Notes in Computer Science*, Springer (July 2011) 194–211
5. Boley, H.: PSOA RuleML: Integrated Object-Relational Data and Rules. In Faber, W., Paschke, A., eds.: *Reasoning Web. Web Logic Rules (RuleML 2015) - 11th Int’l Summer School 2015*, Berlin, Germany, July 31- August 4, 2015, *Tutorial Lectures*. Volume 9203 of *LNCS.*, Springer (2015)
6. Zou, G., Boley, H.: PSOA2Prolog: Object-Relational Rule Interoperation and Implementation by Translation from PSOA RuleML to ISO Prolog. In: *Proc. 9th International Web Rule Symposium (RuleML 2015)*, Berlin, Germany. *Lecture Notes in Computer Science*, Springer (August 2015)
7. Boley, H.: Integrating Positional and Slotted Knowledge on the Semantic Web. *Journal of Emerging Technologies in Web Intelligence* **4**(2) (November 2010) 343–353
8. Berners-Lee, T., Connolly, D., Kagal, L., Scharf, Y., Hendler, J.: N3Logic: A Logical Framework For the World Wide Web. *Theory and Practice of Logic Programming (TPLP)* **8**(3) (May 2008)
9. Kifer, M., Yang, G., Wan, H., Zhao, C.: *ERGO^{Lite}* (a.k.a. *Flora-2*): User’s Manual, v1.1 (2015) <http://flora.sourceforge.net/docs/floraManual.pdf>.
10. Boley, H., Kifer, M.: *RIF Basic Logic Dialect (Working Draft)* (October 2007) W3C Working Draft, <https://www.w3.org/TR/2007/WD-rif-bld-20071030/>.