PSOA RuleML API: a tool for processing abstract and concrete syntaxes
Al Manir, Mohammad Sadnan; Riazanov, Alexandre; Boley, Harold; Baker, Christopher J. O.
PSOA RuleML API: A Tool for Processing Abstract and Concrete Syntaxes

Mohammad Sadnan Al Manir¹, Alexandre Riazanov¹, Harold Boley² and Christopher J.O. Baker¹

¹ Department of Computer Science and Applied Statistics
University of New Brunswick, Saint John, Canada
{sadnan.almanir,bakerc}@unb.ca
alexandre.riazanov@gmail.com
² Information and Communications Technologies
National Research Council Canada
harold.boley@nrc.gc.ca

Abstract. PSOA RuleML is a rule language which introduces positional-slotted, object-applicative terms in generalized rules, permitting relation applications with optional object identifiers and positional or slotted arguments. This paper describes an open-source PSOA RuleML API, whose functionality facilitates factory-based syntactic object creation and manipulation. The API parses an XML-based concrete syntax of PSOA RuleML, creates abstract syntax objects, and uses these objects for translation into a RIF-like presentation syntax. The availability of such an API will benefit PSOA rule-based research and applications.

1 Introduction

F-logic [1] and W3C RIF [2] define objects (frames) separately from functions and predicates. POSL and PSOA RuleML [3, 4] provide an integration of object identifiers with applications of functions or predicates to positional or slotted arguments, called positional-slotted, object-applicative (psoa) term. While RIF requires different kinds of terms for positional and slotted information as well as for frames and class memberships, PSOA RuleML can express them with a single kind of psoa term. As a result, PSOA rules permit a compact way of authoring rule bases, which are as expressive as POSL and semantically defined in the style of RIF-BLD. The constructs of PSOA RuleML are described in [3] in detail. In this paper, ‘psoa’ in lower-case letters refers to a kind of terms while ‘PSOA’ in upper-case letters refers to the language.

Here we describe an open-source PSOA RuleML API. The inspiration comes from well-known APIs for Semantic Web languages such as the OWL API [5] and the Jena API [6]. The existence of these APIs facilitates a lot of experimental research and development in Semantic Web technologies and we hope that our API will have a similar effect on the PSOA adoption. Our API allows creation of objects corresponding to PSOA constructs, such as constants, variables, tuples, slots, atoms, formulas, rules, etc., using factory-based method calls, as well as
traversal of those objects using simple recursive traversal. Moreover, it supports parsing of XML-based PSOA documents and generation of presentation syntax. Thus, users will be able to employ this API for rule processing, including rule authoring, rule translation into other languages, rule-based applications, and rule engines.

Due to space constraints, here we briefly describe the key features of the PSOA RuleML presentation syntax. The language is best described using conditions and rules built over various terms, centered around psoa terms in particular.

We begin with the disjoint sets of alphabets of the language. The alphabets include a countably infinite set of constant and variable symbols, connective symbols (e.g., \texttt{And,Or,:-}), quantifiers (e.g., \texttt{Exists,Forall}), and other auxiliary symbols (e.g., \texttt{=,#,#-,External,Group,(,),<,>,^^,.}).

The language contains literal constants and IRI constants, the latter sometimes abbreviated as short constants.

The following examples illustrate double-type and string-type literal constants:

\begin{verbatim}
"27.98"\texttt{^xs:double} "The New York Times"\texttt{^xs:string}
\end{verbatim}

Constants like \texttt{family}, \texttt{kid} are short constants.

Each variable name is preceded by a ‘?’ sign, such as \texttt{?1,?Hu,?Wi, etc}.

In a psoa term, the function or predicate symbol is instantiated by an object identifier (OID) and applied to zero or more positional or named arguments. The positional arguments are referred to as tuples while named arguments (attribute-value pairs) are called slots.

For example, a psoa term (an atom), containing \texttt{family} relation with the OID \texttt{?inst}, tuples husband \texttt{?Hu}, and wife \texttt{?Wi}, along with a slot child->\texttt{?Ch} can be represented as follows:

\texttt{?inst#family(?Hu ?Wi child->?Ch)}

Terms include psoa terms as well as several different types of logic terms, such as constants and variables, equality, subclass, and external terms.

An atomic formula with \texttt{f} as the predicate is defined as \texttt{f(...) in general. PSOA applies a syntactic transformation to incorporate the OID, which results in the objectified atomic formula \texttt{o#f(...), with o as the OID and f acting as its class. The OID is represented by a stand-alone ‘?’ for a ground (variable-free) fact, an existentially-scoped variable for a non-ground fact or an atomic formula in a rule conclusion, and a stand-alone ‘?’ as an anonymous variable for any other atomic formula.}

Condition formulas are used as queries or rule premises. Conjunction and disjunction of formulas are denoted by \texttt{And and Or}, respectively. Formulas with existentially quantified variables are also condition formulas. An example of a condition formula is given below:

\begin{verbatim}
\texttt{And(?2#married(?Hu ?Wi) Or(?3#kid(?Hu ?Ch) ?4#kid(?Wi ?Ch)))}
\end{verbatim}

Aside from the condition formulas, the premise can also contain atomic formulas and external formulas.
A conclusion contains a head or conjunction of heads. A head refers to an atomic formula which can also be existentially quantified. A conclusion example is given below:

\[ \text{Exists } ?1 \ ( ?1#\text{family}(\text{husb}->?Hu \ \text{wife}->?Wi \ \text{child}->?Ch) ) \]

An implication contains both conclusion and condition formulas. A clause is either an atomic formula or an implication. A rule is generated by a clause within the scope of the \textit{Forall} quantifier or solely by a clause. Several formulas can be collected into a \textit{Group} formula.

The \textit{Group} formula below contains a universally quantified formula, along with two facts. The \textit{Forall} quantifier declares the original universal argument variables as well as the generated universal OID variables ?2, ?3, ?4. The infix \:-\: separates the conclusion from the premise, which derives the existential \textit{family} frame from a \textit{married} relation \textit{And} from a \textit{kid} of the \textit{husb} \textit{Or} \textit{wife}. The following example from [3] shows an objectified form on the right.

\[
\text{Group (}
\text{Forall } ?Hu \ ?Wi \ ?Ch (}
\text{family(husb->?Hu wife->?Wi child->?Ch) :-}
\text{And(married(?Hu ?Wi)} \text{Or(kid(?Hu ?Ch))} )
\text{married(Joe Sue)}
\text{kid(Sue Pete)}
\text{)}
\]

\[
\text{Group (}
\text{Forall } ?Hu \ ?Wi \ ?Ch ?2 ?3 ?4 (}
\text{Exists } ?1 (}
\text{?1#family(husb->?Hu wife->?Wi child->?Ch)) :-}
\text{And(?2#married(?Hu ?Wi)} \text{Or(?3#kid(?Hu ?Ch) ?4#kid(?Wi ?Ch))} )
\text{_1#married(Joe Sue)}
\text{_2#kid(Sue Pete)}
\text{)}
\]

The objectified \textit{family} term in the rule conclusion is slotted with 3 slots:

\[ ?1#\text{family}(\text{husb}->?Hu \ \text{wife}->?Wi \ \text{child}->?Ch) \]

The rules’s condition formulas use the relations \textit{married} and \textit{kid}, containing only 2-tuples Hu, Wi, and Ch:

\[ ?2#\text{married(?Hu ?Wi)} \ ?3#\text{kid(?Hu ?Ch)} \ ?4#\text{kid(?Wi ?Ch)} \]

The next section will describe the API components and their uses. We begin by describing the organization of the package, then illustrate the object creation and traversal as well as parsing the PSOA/XML input, and rendering in presentation syntax. For all of these operations, we use the objectified \textit{family} example above. Finally, we conclude by mentioning the scope of using our API with other complementary tools and potential work directions in the future.

\section{The API Structure and Functionality}

\subsection{Package Organization}

The API is divided into two main components: one is for the creation and traversal of abstract syntax objects and the other is for parsing and rendering of those objects.

The \textit{AbstractSyntax} is the top level class for factories and contains all Java interfaces for different types of abstract syntax objects. A simple implementation of \textit{AbstractSyntax} interfaces is in the \textit{DefaultAbstractSyntax} class, which is suitable for most purposes. However, more demanding uses may require custom implementations of the interfaces.
The package also contains a Parser class, which provides PSOA/XML parsing and translation into presentation syntax. Parsing is implemented using the Java Architecture for XML Binding (JAXB) [7], which creates equivalent Java objects based on XML schema files. The schema is a straightforward encoding of the syntactic construct hierarchy and is available in [8]. The parsed XML is then converted to the abstract syntax by calling factory methods.

Figure 1 presents the most important classes and interfaces implementing different types of syntactic constructs. Each of the names in a rectangular box represents a Java interface, and is kept as close as possible to the presentation syntax construct names. The corresponding implementations of these interfaces use Java inheritance, shown by the solid arrows.

The interface Construct sits at the top of the hierarchy. Group can be populated with more Groups and Rules. Both universal facts and universal rules are represented by the Rule class, which encapsulates a Clause. The interface Clause represents either an implication or an atomic formula. Implication is represented by the interface Implies whereas Atomic represents atomic formulas. The generalized interface Atomic is implemented either by Atom (representing a psoa term like ?2#married(?Hu ?Wi)) or by Subclass (representing a subclass term like student##person) or by Equal (equality term like ?cost = ?spent).

Implementations of the generalized Formula interface represent either disjunction, conjunction, or existential formulas by implementing Formula.Or (represented as Or(?3#kid(?Hu ?Ch) ?4#kid(?Wi ?Ch)), Formula.And (And instead of Or), and Formula.Exists, respectively. In addition to atomic formulas, external formulas (External(func:numeric-add(?cost1 ?cost2))) can be represented using the Formula.External interface. The generalized interface Term is
represented by implementing Const for different kinds of constants, Var for variables like ?Hu, ?Wi, Expr for expressions denoting psoa terms, and External for external Expressions. Constants can be either Const Literal (e.g., "47.5"^xs: float, with Symapce referring to xs:float) or Const Constshort (e.g., family ,kid). Finally, the interface Psoa is implemented to represent objectified functions or predicates with membership symbol ‘#’ and tuples and slots as arguments, e.g., inst#family(Homer Merge child->Bart). The internal nodes Sentence, Formula, Atomic, Term, Const, and Expr, are generalized classes and implemented by more specific classes.

2.2 Construction of Abstract Syntax Objects

The abstract syntax objects are constructed by factory-based createX methods calls, X being the object type name. The rest of this paper represents each method in emphasized font. A factory can be created as follows:

```
DefaultAbstractSyntax absSynFactory = new DefaultAbstractSyntax()
```

We are going to illustrate the creation of facts and rules below.

Construction of Facts A fact is of type Atomic. Let us look at the first fact that tells us Joe and Sue are married to each other with the OID 1, whereas the second fact says Pete is the kid of Sue with the OID 2, each fact referring to a psoa term.

The creation of fact 1#married(Joe Sue) starts by creating the four constants _1, married, Joe, Sue as const_1, const_married, const_Joe, const_Sue, respectively using the method createConst Constshort.

```
Const_Constshort const_1 = absSynFactory.createConst_Constshort("_1")
Const_Constshort const_married = absSynFactory .createConst_Constshort("married")
Const_Constshort const_Joe = absSynFactory .createConst_Constshort("Joe")
Const_Constshort const_Sue = absSynFactory .createConst_Constshort("Sue")
```

Tuples const_Joe and const_Sue are constructed by the method createTuple. The list of such tuples is referred to as a tuplesList.

```
Tuple tuples = absSynFactory.createTuple(tuplesList)
```

Method createPsoa assembles _1, married and tuples into a psoaTerm, while null indicates the absence of slots.

```
Psoa psoaTerm = absSynFactory .createPsoa(const_1, const_married, tuples, null)
```

Here is how we create an atom:

```
Atom atom = absSynFactory.createAtom(psoaTerm)
```

Thus, we use the method createAtom for creating a fact of type Atom, createEqual for a fact of type Equal, and createSubclass for type Subclass. This creation is completed by the createClause and createRule method calls. The representation for creating the fact 2#kid(Sue Pete) is similar to the method calls described above, hence omitted.
Construction of Rules A rule contains condition and conclusion. We will start with the condition formula, which is a conjunction of the atomic formula \(?2#married(?Hu ?Wi)\) and disjunction of two atomic formulas, \(?3#kid(?Hu ?Ch)\) and \(?4#kid(?Wi ?Ch)\).

\[
\quad \exists ?1 ( \\
\quad \quad \text{?1#family(husb->?Hu wife->?Wi child->?Ch)} :- \\
\quad \quad \text{And(?2#married(?Hu ?Wi) Or(?3#kid(?Hu ?Ch) ?4#kid(?Wi ?Ch))))}
\]

The following code snippet creates the disjunction of two atoms. Method `createFormula_Or` defines the disjunction of two atomic formulas, `atomOr_1` and `atomOr_2`. The OIDs \(?3\)` and \(?4\)` (var_4), as well as tuples \(?Hu\)` (var_Hu), \(?Wi\)` (var_Wi) and \(?Ch\)` (var_Ch), are variables. Only the construction of \(?3\)` (var_3) is shown below to avoid repetition (...).

\[
\begin{align*}
\text{Var } & \text{var}_3 = \text{absSynFactory.createVar("3")} \\
\text{...}
\end{align*}
\]

Tuple `tuples` = `absSynFactory.createTuple(tuplesList_1)`

Psoa `psoaTerm_1` = `absSynFactory.createPsoa(var_3, const_kid, tuples, null)`

Atom `atomOr_1` = `absSynFactory.createAtom(psoaTerm_1)`

...Atom `atomOr_2` = `absSynFactory.createAtom(psoaTerm_2)`

Both of the atomic formulas `atomOr_1` and `atomOr_2` are in a list called `formulaOrList`.

\[
\text{Formula_Or} \text{ formula_Or} = \text{absSynFactory.createFormula_Or(formulaOrList)}
\]

The conjunction of the newly created `formula_Or` and another atomic formula `atom_And`, \(?2#married(?Hu ?Wi)\) is described next. Here \(\text{var}_2\), \(\text{var}_\text{Hu}\) and \(\text{var}_\text{Wi}\) denote the variables \(?2\), \(?Hu\) and \(?Wi\), respectively. The code below does using the method `createFormula_And`. The list `formulaAndList` contains atomic formulas `atom_And` and `formula_Or`. The conjunction formula `formula_And` is the rule premise and created as follows:

\[
\text{Formula_And} \text{ formula_And} = \text{absSynFactory.createFormula_And(formulaAndList)}
\]

We now move on to the rule `Head` creation, which is a psoa term containing the OID \(?1\) with `family` class name and three slots, `husb->?Hu`, `wife->?Wi`, and `child->?Ch`, in \(?1#family(husb->?Hu wife->?Wi child->?Ch)\) as arguments. These slots will be called `slot_1`, `slot_2`, and `slot_3`, respectively.

\[
\text{Var } \text{var}_1 = \text{absSynFactory.createVar("1")} \\
\text{...}
\]

Slot `slot_1` = `absSynFactory.createSlot(const_husb, var_Hu)`

Slot `slot_2` = `absSynFactory.createSlot(const_wife, var_Wi)`

Slot `slot_3` = `absSynFactory.createSlot(const_child, var_Ch)`

The list of slots `slot_1`, `slot_2` and `slot_3`, called `slotsList`, is used to create the psoa term.

\[
\text{Psoa } \text{psoa} = \text{absSynFactory.createPsoa(var}_1, \text{const_family, null, slotsList})
\]
The atom created is \texttt{atom.head}, and thus the rule head is created by the method \texttt{createHead}, where the variable \texttt{var.1} is existentially quantified. Both existentially and universally quantified variables are treated as a list of variables, called \texttt{varsList}.

Head rule head = absSynFactory.createHead(varsList, atom.head)

The method \texttt{createImplies} combines the rule head, \texttt{rule.head} and the rule premise, \texttt{formula_And}, into an implication. Method \texttt{createClause} creates the implication. Finally, method \texttt{createRule} collects all the universally quantified variables, \texttt{var.Hu}, \texttt{var.Wi}, \texttt{var.Ch}, \texttt{var.2}, \texttt{var.3}, and \texttt{var.4} into a \texttt{varsList} and creates the rule with the clause.

Implies implication = absSynFactory.createImplies(rule_head, formula_And)
Clause clause = absSynFactory.createClause(implication)
Rule rule = absSynFactory.createRule(varsList, clause)

2.3 Abstract Syntax Structure Traversal

Our implementation recursively traverses the object tree generated from the abstract syntax structure and is usually simpler than writing visit methods [9] as used in OWL API.

All components of the abstract syntax structure can be accessed directly by the corresponding accessor methods, which are \texttt{getX} methods. The generalized classes (see Figure 1) are \texttt{Sentence}, \texttt{Formula}, \texttt{Atomic}, \texttt{Term}, \texttt{Const}, and \texttt{Expr}, containing \texttt{isX} methods to recognize the specific instance types. Alternatively, specific classes of particular instances have to be identified, e.g., by using \texttt{instanceof}.

For an atomic formula, an \texttt{isX} method in \texttt{Atomic} class needs to recognize if the instance is of type \texttt{Atom}, \texttt{Subclass}, or \texttt{Equal} object. This principle applies to each of the generalized classes.

For example, \texttt{isEqual} method in generalized class \texttt{Atomic} recognizes the instance of Equality atom \texttt{?cost = "47.5"^^xs:float}. Immediately, a cast is made as the instance type \texttt{Equal} and \texttt{getLeft} and \texttt{getRight} methods are called, each referring to an instance of another generalized class \texttt{Term}. Class \texttt{Term} contains appropriate \texttt{isX} methods, which use similar techniques to find out if the instance is of type \texttt{Const}, or \texttt{Var}, or an \texttt{External} expression.

assert this instanceof AbstractSyntax.Equal
return (AbstractSyntax.Equal) this

AbstractSyntax.Term getLeft()
AbstractSyntax.Term getRight()

Method \texttt{getLeft}, in this case, retrieves the instance of \texttt{Var} and thus string variable \texttt{?cost} is retrieved by the method \texttt{getName} as the variable instance. On the other hand, \texttt{getRight} refers to the instance of type \texttt{Const}. Method \texttt{isConstLiteral} recognizes \texttt{ConstLiteral} involving the literal and the type \texttt{float} involving the instance of type \texttt{SymSpace}. The literal object 47.5 is retrieved as string by the method \texttt{getLiteral}. Finally, \texttt{xs:float} object is retrieved by the method \texttt{getValue} as an instance of type \texttt{SymSpace}.
Thus, the traversal of objects in the API structure follows the same strategy of going down to most specific instances in a recursive manner for both facts and rules.

2.4 Parsing and Rendering

Aside from creating and traversing objects, the API is able to parse PSOA/XML inputs and render them in human readable presentation syntax.

In section 2.1 we discuss an XML schema for PSOA RuleML. We generate the XML parser with the help of JAXB, which creates Java classes from a schema traversal, where the ultimate output of the parser is abstract syntax objects.

The following example shows a transformation of an XML input for a fact and its rendering in presentation syntax.

```xml
<Atom>
    <Member>
        <instance>
            <Const type="&psoa;iri">inst1</Const>
        </instance>
        <class>
            <Const type="&psoa;iri">family</Const>
        </class>
    </Member>
    <tuple>
        <Const type="&psoa;iri">Joe</Const>
        <Const type="&psoa;iri">Sue</Const>
    </tuple>
    <slot>
        <Const type="&psoa;iri">Child</Const>
        <Const type="&psoa;iri">Pete</Const>
    </slot>
</Atom>

inst1#family(Joe Sue Child->Pete)
```

A `toString` method in each class implements this pretty-printing, which follows the same traversal procedure described in section 2.3.

3 Conclusion and Future Work

The API is open-source and hosted in [10]. The companion effort PSOA2TPTP [11] has developed a reference translator for PSOA RuleML, which facilitates inferring using TPTP reasoners (see e.g., [12]). One component of the translator is a parser for the presentation syntax. Our API will greatly benefit from including this presentation syntax parser. The other component of the PSOA2TPTP translator is its mapping from abstract syntax objects to TPTP. Combined with our API, this will also make PSOA/XML executable on the TPTP-aware VampirePrime [13] reasoner.

Currently, the API can render PSOA/XML only into presentation syntax. As an extension, we plan to also include the translation of abstract syntax objects back to PSOA/XML.

We have been using the API in our HAiku work [14], where PSOA is used to capture semantic modeling of relational data and needed, at least, to support
authoring, including syntactic and, to some extent, logical validation (consistency checking). We also plan to use it for automatic generation of Semantic Web services from declarative descriptions.

PSOA RuleML API has become an input to the Object Management Group’s API4KB effort [15], which tries to create a universal API for knowledge bases that among other things combines the querying of RDF-style resource descriptions, ODM/OWL2-style ontologies, and RIF RuleML-style rules.

References

12. System on TPTP. http://www.cs.miami.edu/~tptp/cgi-bin/SystemOnTPTP.
Publishing Agreement

The Author is an employee of the National Research Council Canada (NRC) and is therefore publishing under Crown Copyright. The copyright to this article is retained by the Crown of Canada. The NRC grants to Springer Science + Business Media B.V., Dordrecht for the article identified below the non-exclusive rights to reproduce and distribute the article, including reprints, translations, photographic reproductions, microform, electronic form (offline, online) or any other reproductions of similar nature. Springer will take any necessary steps to protect the copyright of the Crown against infringement by third parties. It will have the copyright notice inserted into all editions of the article according to the provisions of the Universal Copyright Convention (UCC) and dutifully take care of all formalities in this connection. The copyright notice will appear in the following form: Copyright © Her Majesty the Queen in Right of Canada 2011.

The NRC warrants that this contribution is original.

An author may self-archive an author-created version of his/her article on his/her own website and or in his/her institutional repository. He/she may also deposit this version on his/her funder’s or funder’s designated repository at the funder’s request or as a result of a legal obligation, provided it is not made publicly available until 12 months after official publication. He/she may not use the publisher’s PDF version, which is posted on www.springerlink.com, for the purposes of self-archiving or deposit. Furthermore, the author may only post his/her version provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer’s website. The link must be accompanied by the following text: “The original publication is available at www.springerlink.com”.

Please use the appropriate DOI for the article. Articles disseminated via www.springerlink.com are indexed, abstracted, and references by many abstracting and indexing services, bibliographic networks, subscription agencies, library networks, and consortia.

After submission of this agreement signed by NRC, changes of authorship or in the order of the authors listed will not be accepted by Springer.


Title of the article: PSOA RuleML_API: A Tool for Processing Abstract and Concrete Syntaxes

Author(s):
Mohammad Sadnan Al Manir, Alexandre Riazanov, Harold Boley and Christopher J.O. Baker

E-Mail of Author(s)
sadnan.almanir@unb.ca, bakerc@unb.ca
alexandre.riazanov@gmail.com
harold.boley@nrc.gc.ca

Signed at Fredericton, New Brunswick on JUN 27 2012
(Date)

NATIONAL RESEARCH COUNCIL OF CANADA

Per: Andrew Reddick, Director Research
INSTITUTE FOR INFORMATION TECHNOLOGY
COPYRIGHT RELEASE FORM AND IP CONTENT
FORMULAIRE DE CESSION DES DROITS D'AUTEUR ET CONTENU DE PI

To / À: Andrew Reddick, Director Research
From / De: Stephen MacKay

Title: RSOA.RuleML API: A Tool for Processing Abstract and Concrete Syntaxes
Titre: 

Author(s): Mohammad Sadnan Al Manir, Alexandre Riazanov, Harold Boley and Christopher J.O. Baker
Auteur(s): 

Journal, Conference, Workshop / journal, conférence, atelier:
Conference Proceedings: RuleML 2012 - The 6th International Symposium on Rules

Planned Publication Date:
Date prévue de publication: August 2012

Attached is an article that I have reviewed and am recommending for publication. I have consulted with the author(s) and our Business Development Office as appropriate to ensure that the BD Office is aware of the technology and has had the opportunity to make recommendations on IP protection and technology transfer. The IP protection & technology transfer aspects have been considered within the context of related NRC-IIT technology and existing and potential licenses.

I have verified that the publication procedure has been followed in filling out the selected form, and that to the best of my knowledge, we are granting the publisher only those rights that are necessary to have the work published. I believe that publication of this document will enhance the reputation and influence of NRC-IIT and the authors in the knowledge commons.

19 - Copyright Licence / Licence de droit d'auteur
22 - Licence to Publish / Droit de publier (IEEE)
21 - Copyright Transfer
22 - Publication Agreement / Entente de publication (SPRINGER)
22 - Licence to Publish / Droit de publier

(Stephen MacKay)

Internet Logic / Logique Internet

JUN 27 2012

NRC-IIT, March 2008 / ITI-CNRC mars 2008