On the Implementation of an Algebra of Lightweight Ontologies

Rômulo C. Magalhães  
Department de Informática,  
PUC-Rio,  
Rio de Janeiro, Brasil  
rmagalhaes@inf.puc-rio.br

Marco A. Casanova  
Department de Informática,  
PUC-Rio,  
Rio de Janeiro, Brasil  
casanova@inf.puc-rio.br

Bernardo P. Nunes  
PUC-Rio,  
UNIRIO  
Rio de Janeiro, Brasil  
bernardo@cceed.puc-rio.br

Giseli Rabello Lopes  
Departamento de Ciência da Computação,  
UFRJ,  
Rio de Janeiro, Brasil  
giseli@dcc.ufrj.br

ABSTRACT
This paper first argues that ontology design may benefit from treating ontologies as theories and from the definition of a set of operations that map ontologies into ontologies, especially their constraints. The paper then defines the class of ontologies used and proposes four operations to manipulate them. It proceeds to discuss how the operations may help design new ontologies. The core of the paper describes an implementation of the operations as a Protégé plug-in, called OntologyManagerTab, and includes use case examples to validate the discussion.

CCS CONCEPTS
Data Design, Evolution and Migration; Data Integration; Data Models; Semantic Web and Databases.

KEYWORDS
Constraint Definition; Ontology Design; Linked Data; Protégé; OntologyManagerTab.

1 INTRODUCTION
The Linked Data principles [4][5] suggest a way to publish databases on the Web that facilitates interoperability. These principles emphasize the definition of the conceptual structure of the data through the reuse of known ontologies, thus minimizing the need for alignment between conceptual schemas, a difficult and error prone task, which lies at the core of the interoperability issue. We argued elsewhere that certain familiar ontology design problems are profitably addressed by treating ontologies as theories and by defining a set of operations on ontologies [7][9].

Briefly, we define an ontology as a pair $O=(V, \Sigma)$ such that $V$ is a vocabulary and $\Sigma$ is a set of constraints in $V$. The theory of $\Sigma$ is the set of all constraints that are logical consequences of $\Sigma$. We emphasize that the constraints in $\Sigma$ capture the semantics of the terms in $V$ and must, therefore, be brought to the foreground. The theory of $\Sigma$ identifies the constraints that are implicitly defined, but which must be considered when using the ontology. The operations we propose create new ontologies out of other ontologies taking into account their constraints, and facilitate the domain specialist to uphold the Linked Data principles. Such operations extend the idea of namespaces to consider constraints and help address familiar ontology design problems, which we now outline.

Consider first the problem of designing an ontology to publish data on the Web. If the designer follows the Linked Data principles, he must select known ontologies, as much as possible, to organize the data so that applications “can dereference the URIs that identify vocabulary terms in order to find their definition”. We argue that the designer should go further and analyze the constraints of the ontologies from which he is drawing the terms to construct his vocabulary. Furthermore, he should publish the data so that the original semantics of the terms is preserved. To facilitate ontology design from this perspective, we introduce three operations on ontologies, called projection, union and deprecation.

Consider now the problem of comparing the expressive power of two ontologies, $O_1=(V_1, \Sigma_1)$ and $O_2=(V_2, \Sigma_2)$. If the designer wants to know what they have in common, he should create a mapping between their vocabularies and detect which constraints hold in both ontologies, after the terms are appropriately mapped. The intersection operation answers this question. We argued elsewhere [6] that intersection is also useful to address the design of mediated schemas that combine export schemas in a way that the data exposed by the mediator is always consistent.

On the other hand, if the designer wants to know what holds in $O_1=(V_1, \Sigma_1)$, but not in $O_2=(V_2, \Sigma_2)$, he should again create a mapping between their vocabularies and detect which constraints hold in the theory of $\Sigma_1$, but not in the theory of $\Sigma_2$. We argued elsewhere that certain familiar
after the terms are appropriately mapped. The difference operation answers this question. Likewise, if the user wants to analyze what changed from one version of an ontology to the other, he should also use the difference operation.

The core of this paper describes an implementation of projection, union, intersection and difference as a Protégé plugin, called OntologyManagerTab, to assist the domain specialist in the process of managing ontologies for data publication, it includes use case examples to validate the discussion. The machinery to handle constraints developed in [6][7][8] and [9] provides the theoretical foundations of the paper. Previous work by the authors [7] introduced the notion of open fragment, which is captured by projection. The design of mediated schemas was addressed in [6]. A preliminary implementation of the operations was described in [16].

The paper is organized as follows. Section 2 reviews background concepts. Section 3 introduces the operations. Sections 4 summarizes related work. Section 5 and 6 discuss the implementation of the operations and present examples. Section 7 contains the conclusions and future work.

2 BACKGROUND

We adopt the basic notion of Description Logic [3]. Very briefly, a vocabulary \( V \) consists of a set of atomic concepts, a set of atomic roles, and the bottom concept \( \bot \). A language in \( V \) is a set of strings, using symbols in \( V \), defining the set of concept descriptions in \( V \) and the set of role descriptions in \( V \). An inclusion in \( V \) is a string of the form \( u \subseteq v \), where \( u \) and \( v \) both are concept descriptions in \( V \) or both are role descriptions in \( V \). Table 1 shows common types of inclusions.

An ontology is a pair \( O=(V, \Sigma) \) such that \( V \) is a vocabulary and \( \Sigma \) is a set of inclusions in \( V \), called the ontology constraints. The theory of \( \Sigma \) (or the theory of \( O \)), denoted \( \tau[\Sigma] \) (or \( \tau[O] \)), is the set of all logical consequences of \( \Sigma \).

We say that two sets of inclusions, \( \Gamma \) and \( \Theta \), are equivalent, denoted \( \Gamma \equiv \Theta \), iff their theories are equal, that is, the set of all logical consequences of \( \Gamma \) is equal to that of \( \Theta \). Likewise, two ontologies \( O_1=(V_1, \Sigma_1) \) and \( O_2=(V_2, \Sigma_2) \) are equivalent, also denoted \( O_1 \equiv O_2 \), iff \( \Sigma_1 \) and \( \Sigma_2 \) are equivalent.

We adopt in this paper the family of lightweight languages, which is equivalent to the family \( DL-Lite^\mathbb{Y}_{core} \) [2] and the proof procedure based on constraint graphs developed in [6][7][8] and [9]. Table 1 lists the constraints allowed in the lightweight family.

3 OPERATIONS ON ONTOLOGIES

Definition 1 introduces the ontology operations we propose. It is important to highlight that the new ontology, obtained from the execution of these operations, presents a set of constraints that considers the semantics of the constraints of the ontologies involved.

<table>
<thead>
<tr>
<th>Name</th>
<th>Inclusion type</th>
<th>Informal semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Constraint</td>
<td>(≥ 1 P) ⊆ C</td>
<td>property P has class C as domain, that is, if (a,b) is a pair in P, then a is an individual in C.</td>
</tr>
<tr>
<td>Range Constraint</td>
<td>(≥ 1 P) ⊆ C</td>
<td>property P has class C as range, that is, if (a,b) is a pair in P, then b is an individual in C.</td>
</tr>
<tr>
<td>minCardinality Constraint</td>
<td>C ∩ (≥ k P) or</td>
<td>property P or its inverse P' maps each individual in class C to at least k distinct individuals.</td>
</tr>
<tr>
<td>maxCardinality Constraint</td>
<td>C ∩ (≥ K+1 P) or</td>
<td>property P or its inverse P' maps each individual in class C to at most k distinct individuals.</td>
</tr>
<tr>
<td>Subset Constraint</td>
<td>C ⊆ D</td>
<td>each individual in C is also in D, that is, class C denotes a subset of class D.</td>
</tr>
<tr>
<td>Disjointness Constraint</td>
<td>C ⊈ D</td>
<td>no individual is in both C and D, that is, classes C and D are disjoint.</td>
</tr>
</tbody>
</table>

Table 1. Common inclusion types used in conceptual modeling.

Definition 1: Let \( O_1=(V_1, \Sigma_1) \) and \( O_2=(V_2, \Sigma_2) \) be two ontologies, \( W \) be a subset of \( V_b \), and \( \Psi \) be a set of inclusions in \( V_f \).
1. The projection of \( O_1=(V_1, \Sigma_1) \) over \( W \), denoted \( \tau[\Psi](O_1) \), returns the ontology \( O_2=(V_2, \Sigma_2) \), where \( V_2=W \) and \( \Sigma_2 \) is the set of constraints in \( \tau[\Sigma_1] \) that use only classes and properties in \( W \).
2. The deprecation of \( \Psi \) from \( O_1=(V_1, \Sigma_1) \), denoted \( \delta[\Psi](O_1) \), returns the ontology \( O_2=(V_1, \Sigma_2) \), where \( V_2=V_1 \) and \( \Sigma_2=\Sigma_1-\Psi \).
3. The union of \( O_1=(V_1, \Sigma_1) \) and \( O_2=(V_2, \Sigma_2) \), denoted \( O_1 \cup O_2 \), returns the ontology \( O_3=(V_3, \Sigma_3) \), where \( V_3=V_1 \cup V_2 \) and \( \Sigma_3=\Sigma_1 \cup \Sigma_2 \).
4. The intersection of \( O_1=(V_1, \Sigma_1) \) and \( O_2=(V_2, \Sigma_2) \), denoted \( O_1 \cap O_2 \), returns the ontology \( O_3=(V_3, \Sigma_3) \), where \( V_3=V_1 \cap V_2 \) and \( \Sigma_3=\Sigma_1 \cap \Sigma_2 \).
5. The difference of \( O_1=(V_1, \Sigma_1) \) and \( O_2=(V_2, \Sigma_2) \), denoted \( O_1- O_2 \), returns the ontology \( O_3=(V_3, \Sigma_3) \), where \( V_3=V_1 \) and \( \Sigma_3=\Sigma_1-\tau[\Psi] \).

Note that deprecation does not reduce to difference since, in general, we have \( \tau[\Sigma_1]-\tau[\Psi] \neq \tau[\Sigma_1]-\tau[\Psi] \).

We also note that the ontology \( O_3 \) that results from an operation is unique, by definition. However, there might be several ontologies that are equivalent to \( O_3 \). For example, if \( O_2=(V_2, \Sigma_2) \) is the projection of \( O_1 \) on \( W \), there might be several sets of constraints that are equivalent to the set of constraints in the theory of \( O_1 \) that use only terms in \( W \).

The projection operation allows the designer to define a set \( W \) containing just a few terms from the vocabulary of an ontology. This set retains the semantics of the terms from the original vocabulary in \( W \) through the constraints derived from those of the original ontology. It is also the only operation among those implemented in this dissertation that has a single ontology as argument. The main advantage of the Projection procedure is to automate the onerous task of the do- main specialist in the
formalization of new ontologies by extracting the needed concepts and their dependencies, allowing the reuse of widely consolidated terms with little work.

The applications that adopt the principles of Linked Data have the challenge of providing its users with integrated information from multiple data sources that may or may not contain overlapping data, hence the importance of the Union and Intersection operations.

Usually, the process of integrating multiple ontologies consists in the union of two versions of ontologies unbeknownst whether there was a common ontology that originated them. Also, the union between data from distinct domains may generate conflicts and inconsistencies depending on the versions of the ontologies used by each domain.

Similarly to the Union operation, the domain specialist may want to extract only the overlapping information while consulting multiple data sources. This is achieved by the Intersection operation.

Ontologies evolve over time due to changes in the domain they represent or due to the fact that they have been built in a collaborative way and, therefore, need to be updated to represent a common understanding to different users. To detect modifications between two versions of the same ontology, we have the Difference operation, that compares two ontologies and returns the terms and constraints that are present in the first, but not in the second.

According to (M. Klein, 2004), the differences between two ontologies can be classified as simple or complex. The simple differences are those that do not affect the structure of the ontology, such as the change of names of classes, properties or data types. On the other hand, complex differences are those that affect the ontology structure, include modifications in the class hierarchy or in constraints, such as disjunction and cardinality restrictions. The Difference operation addresses both types of differences.

4 RELATED WORK

Reference [11] argues that the Linked Open Data (LoD) Cloud, in its current form, is only of limited value for furthering the Semantic Web vision. They discuss that the Linked Open Data Cloud can be transformed from “merely more data” to “semantically linked data” by overcoming problems such as lack of conceptual descriptions for the datasets, schema heterogeneity and absence of schema level links. Along this line, we advocated that the design of Linked Data sources must include constraints derived from those of the underlying ontologies.

We note that the problem we cover in this paper cannot be reduced to a question of ontology alignment in the context of Linked Data, addressed for example in [17], [21], [10] and [15]. Indeed, we stress that the problem we focus on refers to bootstrapping a new ontology (including its constraints) through the implementation of ontology algebra operations over one or more existing ontologies.

Some tools, such as Prompt [14] and ODEMerge [18], allow the user to combine two or more ontologies in a semiautomatic or in an automatic way. Other tools, such as PromptDiff [13] and OntoDiff [19], deal with ontology change detection. Our tool offers a complete environment to design and maintain ontologies, which allows applying a series of operations over one or more ontologies and enabling reuse, versioning, evolution and integration of ontologies.

Reference [20] proposes a tool that implements the projection operation by the creation of a database view resulting from query execution. However, this tool does not allow the generation of semantic information captured by the constraints that apply to the vocabulary terms.

Other tools, such as OAPT, proposed in [1], address the question of partitioning an ontology into modules to facilitate ontology reuse. The modularization operation is akin to our projection operation. However, we go further and consider other operations that, in the end, show how to combine modules into larger ontologies. Finally, previous work by the authors [7] introduced the notion of open fragment, captured by the projection operation, and [9] covered some of the operations discussed in this paper.

5 IMPLEMENTATION OF THE OPERATIONS

The OntologyManagerTab offers the ontology operations described in Section 3, integrated with traditional ontology management features. OntologyManagerTab was developed in Java as a tab plug-in of Protégé 3.4.8 (this implementation might require minor modifications to work with other versions of Protégé). However, OntologyManagerTab works in a completely independent manner from the main framework, using Protégé only as a Graphical User Interface (GUI) enclosure. In other words, all the functionalities provided by OntologyManagerTab do not rely on any of the Protégé libraries, making the tool easily adaptable as a plug-in for any other framework or as a stand-alone software.

OntologyManagerTab performs a procedure that loads only the lightweight kernel of an ontology, that is, the lightweight constraints of the ontology, according to the DL-Lite\(_{\text{core}}\) family. Thus, it guarantees that the operations will be executed only over lightweight constraints, consistently with the use of constraint graphs.

The software source code is divided into three packages: Application, Ontology and Main. The Application package contains classes that implement the Java interfaces required for the application to run. The Main package contains the OntologyManagerTab class, that implements the discussed operations and integrates the software with the Protégé framework. Finally, the Ontology package contains the classes that implement constraint graphs, as well as the procedures to load the lightweight kernel of ontologies and save the results obtained.

6 EXAMPLES
To exemplify a well-known ontology that contained some of the needed terms, the domain specialist discovered certain terms and constraints to keep, since they are equivalent. Experiments Processing Times in seconds.

<table>
<thead>
<tr>
<th>Sample</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parliament Ontology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Load</td>
<td>0.724</td>
<td>1.261</td>
<td>1.130</td>
<td>0.941</td>
<td>0.764</td>
<td>0.984</td>
<td>0.689</td>
<td>0.436</td>
<td>0.589</td>
<td>0.779</td>
<td>0.829</td>
</tr>
<tr>
<td><strong>Relationship Ontology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Load</td>
<td>0.955</td>
<td>0.728</td>
<td>0.649</td>
<td>0.883</td>
<td>0.823</td>
<td>0.792</td>
<td>0.589</td>
<td>0.743</td>
<td>1.219</td>
<td>0.509</td>
<td>0.789</td>
</tr>
<tr>
<td><strong>Parliament and Relationship Ontologies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Union</td>
<td>0.428</td>
<td>0.473</td>
<td>0.435</td>
<td>0.479</td>
<td>0.601</td>
<td>0.416</td>
<td>0.488</td>
<td>0.506</td>
<td>0.512</td>
<td>0.476</td>
<td>0.481</td>
</tr>
<tr>
<td>Intersection</td>
<td>0.402</td>
<td>0.483</td>
<td>0.488</td>
<td>0.530</td>
<td>0.471</td>
<td>0.421</td>
<td>0.521</td>
<td>0.490</td>
<td>0.472</td>
<td>0.497</td>
<td>0.477</td>
</tr>
</tbody>
</table>

This section summarizes experiments with the operations. We refer the reader to [12] for detailed experiments with full ontologies from several domains, as well as consistent fragments of these ontologies and a few ontologies created from scratch to test the base cases. All experiments were performed in a machine with a 2 GHz Intel Core 2 Duo processor, 8 GB 1067 MHz DDR3 of memory and using the Java Virtual Machine default configuration.

Table 2 at the end of the paper illustrates the projection operation. The upper left table in Table 2 shows a representation of the input ontology $O_1$, where each lightweight term used in a constraint of $O_1$ appears at least once in the first column of the table and each constraint $e \subseteq f$ of $O_1$ appears in a separate row, with $e$ in column 1 and $f$ in column 2. The upper right table allows the user to select the set of symbols $W$ of the vocabulary to define the projection. The lower left table in Table 2 shows the results of the projection $O_\text{proj} = (V_\text{proj}, \Sigma_\text{proj})$, whereas in the first table each constraint $e \subseteq f$ appears in a separate row, with $e$ in column 1 and $f$ in column 2, as well as its inverse $f \subseteq e$. For example, both Person $\subseteq \neg$Organization and Organization $\subseteq \neg$Person are shown in the lower left table; it is up to the user to decide which constraint to keep, since they are equivalent.

To execute the projection operation, the user first selects “Projection” on the dropdown list that specifies which operation will be executed. Next, the user loads the ontology from which he intends to extract the lightweight kernel by clicking the “Load Ontology Projection” button and selecting the base ontology file. At this point, the base ontology will be displayed in the table on the upper left and its classes and properties will be displayed in the table on the upper right. From this last table, the user then selects the terms to be used for the projection and clicks the “Run” button next to the operation dropdown list. The result of the projection will be displayed in the “Resulting Ontology” table on the lower left. To execute chained procedures, such as $\text{Union}(\text{Intersection}(O_1, O_2), O_3)$, the user must save the ontology resulting from the first operation, $O_1 = \text{Intersection}(O_1, O_2)$, and reload it to execute the second operation, $\text{Union}(O_1, O_3)$.

To better illustrate the usefulness of the application, consider the following Scenario 1: a domain specialist has created an ontology in which he defined certain terms and constraints that he could not find in any related ontology; later on, he discovered a well-known ontology that contained some of the needed terms. To exemplify Scenario 1, we will use the Parliament Ontology, created to specify the organization of the Brazilian government as well as the relationship between politicians and members of the executive branch, as the first ontology created by the domain specialist. Its latest version can be obtained at [22] and a version with the latest data regarding the Brazilian 55th legislature of 2015 at [23]. We adopt the Relationship ontology, which contains a vocabulary for describing relationships between people, as the second ontology. Its latest version can be found at [24].

Using the OntologyManagerTab, the user can specify the type of operation that will be performed. When any operation, except Projection, is selected, the upper right table changes into a copy of the upper left table as shown Table 2. Also, when selecting these other operations, a checkbox and a spinner appear above the upper right table, giving the user the option of using word approximation. This can be done by selecting the checkbox and specifying the number of different characters to be accepted in the comparison; it will ignore blank spaces as well as disregard capitalization of letters and only count the different characters until the specified threshold. The recommended value for the word approximation threshold is 1, just to consider plurals and typos committed by the ontology designer. If the specified threshold is too big, the program may find matches that do not describe the same term, especially if the ontologies in question have classes or properties with small names.

Considering Scenario 1, the user can execute the Intersection between the Parliament and Relationship ontologies using the word approximation option to make sure there is a correlation between them, that is, if their intersection is not empty. After confirming their correlation, it is possible to use the Union between them, as shown in Table 2 to aggregate the rest of the relationships declared in the Relationship ontology to the Parliament ontology and save the resulting new ontology.

Another possibility would be that, after confirming the correlation between the two ontologies, the domain specialist would like to add only a few extra properties and classes from the Relationship ontology. To achieve this, the user would need to load the Relationship ontology and run a Projection over it to select the desired items, remembering to also include the classes and properties that are common between both ontologies, save the resulting ontology and afterwards run a Union between this saved ontology and the Parliament ontology, using the word approximation option with the same configuration as that used in the intersection.
Table 2 shows 10 processing time samples obtained for the operations of Loading, Intersection and Union executed in Scenario 1 with word approximation threshold equal to 1. All measurements are in seconds and the last column shows the average. Still regarding processing times, the projections of approximately 15 or less classes and properties is almost instantaneous with an order of magnitude of 0.02 seconds. All these processing times were measured adding timestamps to the original code.

The only applications similar enough to ours, for comparison, would be Prompt [14] and ODEMerge [18], which, as said in Section 4 allow the user to combine two or more ontologies. However, we were not able to add the same timestamps in order to have the correct metric for time correlation between them and OntologyManagerTab. That aside, on average they took close to 1.5 seconds to do the automatic part of the merge, which would be our Union.

As for the Difference operation provided by OntologyManagerTab, one may use it to compare versions of the same ontology to identify the updates made. This operation can also be used as an intermediate step at the domain specialist discretion. For instance, to perform Deprecations, the user may first perform a projection of the terms he wants to deprecate from the ontology and save the resulting ontology; afterwards he may run the Difference between the first ontology and the saved ontology. The processing times of the Difference operation are similar to those of the Intersection operation, and will vary according to the involved ontologies.

7 CONCLUSION AND FUTURE WORK

Few tools assist the domain specialist in the development of a new ontology that represents a correct understanding of the semantics of the ontologies involved. To address this issue, it is necessary to take into account not only the terms of the original ontologies but also their constraints. This is possible by considering ontologies as logical theories, composed of vocabularies and constraints, and defining the algebraic operations over them.

In this paper, we argued that ontology design may benefit from treating ontologies as theories and from the definition of a set of operations that map ontologies into ontologies, especially their constraints. The core of the paper outlined an implementation of the operations as a Protégé plug-in, the OntologyManagerTab, and included some examples to illustrate the discussion. We refer the reader to [12] for further examples.

As for future work, we intend to extend the tool to more expressive classes of ontologies, using the results in [8] and to improve the GUI with complementary visual representation.

Acknowledgments. This work was partly funded by CNPq under grants 442338/2014-7 and 303332/2013-1, and by FAPERJ, under grant E-26/201.337/2014.

REFERENCES


Referenced Ontologies:

[22] Parliament Ontology owl description file can be found at:
http://www.inf.puc-rio.br/~rmagalhaes/projetos/ParliamentOntology.owl

[23] Parliament Ontology rdf file with data can be found at:
http://www.inf.puc-rio.br/~rmagalhaes/projetos/ParliamentOntology.rdf

[24] Relationship Ontology latet rdf file can be found at:
http://vocab.org/relationship/

Figure 1 - Projection of some classes of the FOAF ontology.
Figure 2 - Union of Parliament and Relationship ontologies.