CEUR Workshop Proceedings
http://ceur-ws.org
ISSN 1613-0073

Vol-596
urn:nbn:de:0074-596-3

ORES-2010
Ontology Repositories and Editors for the Semantic Web

Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web

Hersonissos, Crete, Greece, May 31st, 2010.

Edited by

Mathieu d'Aquin, **The Open University, UK**
Alexander García Castro, **Universität Bremen, Germany**
Christoph Lange, **Jacobs University Bremen, Germany**
Kim Viljanen, **Aalto University, Helsinki, Finland**

10-Jun-2010: submitted by Christoph Lange
11-Jun-2010: published on CEUR-WS.org

Using extended metadata model OMV and metrics in OntoLP Portal

Anderson Bestteti, Larissa A. de Freitas, and Renata Vieira

College Computer, PPGCC,
Ipiranga Avenue, 6681, Partenon, Building 32, Porto Alegre, RS, Brazil
{anderson.bestteti,larissa.freitas,renata.vieira}@pucrs.br
http://www.inf.pucrs.br/~linatural/

Abstract. This paper describes the application of OMV-R, an extension of the OMV metadata model for describing ontologies. The motivation is to help users searching for ontologies and ontology research related resources for reuse. In our extension we included information about ontology evaluation and evolution and we also propose more elaborated ways for describing ontology metrics. The extended model was applied to a web portal for Portuguese ontologies, OntoLP. The model was implemented and evaluated through questionnaires answered by users, researchers and ontology engineers.

Key words: Metamodel, Metrics, Ontology Repository, Ontology Reuse

1 Introduction

One of the main features of ontologies is to enable humans and machines to communicate through semantic formalization. With this purpose, ontologies are used in many areas of Computer Science, such as artificial intelligence, database and software engineering. It has been also the key component for information systems in a increasing variety of domains, well-known examples are the areas of biology, geography and law.

The Semantic Web Project contributed for popularization of ontology as an artifact to build the knowledge representation, shared and reused. With the reuse of ontologies costs can be reduced, mainly in development phase.

As presented in [9], the reuse of ontologies can be seen from two main points: one is to create, extend, specialize and adapt an ontology to build a new one; the other is to combine different ontologies in a single one.

A first problem for reuse is ontology search. Usually, repositories offer poor navigation interfaces [2]. The lack of good quality documentation about available ontologies also contributes to this difficulty. The Metadata models that describe ontologies have been proposed to deal with these problems[6]. The use of a standard vocabulary for ontologies is understood as a basic requisite for a good description of ontologies on the Web.

In this paper we extend and evaluate the previously proposed OMV metadata model through its application to a portal for Portuguese ontologies. The motivation is to help users searching for ontologies and ontology research related resources.

The paper is organized as follows: in Section 2 we present our extension which we call OMV-R; in Section 3 we describe a set of ontology metrics that we use in the description of the ontologies in our portal; in Section 4 we discuss the evaluation of the model; finally, in Section 5 we show the conclusion and future works.

2 The extended metadata model OMV-R

Our extension of the OMV model and its application to a portal aims to assist users in the process of finding out ontologies for reuse.

The original model proposed by [8] does not include information about ontology evaluation and evolution. We included those elements in the model and we also propose more elaborated ways for describing ontology metrics.

According to [13], metrics are useful to help evaluate ontologies during the build and application phase, enabling a quick and simple understanding about what are being modeled through these structures and facilitating the control of its future evolution. This also holds for ontology reuse.

In [10] a study is performed about the maturity of the current process, methodologies and tools focused on ontology reuse. Their strengths and requirements are identified.

Specifically on the evaluation of ontologies, [16] presents a methodology called Requirements-Oriented Methodology for Evaluation Ontologies (ROMEO). This methodology propose a set of criteria for ontology evaluation. Inspired by these previous works, we considered that ontology evaluation is another important feature in the ontology selection process.

As OMV does not provide such information so important to enable reuse, we extended it with new classes as shown in Figure 1.

2.1 Description of the new classes, properties and relations

This section will present the classes, properties and relations such as were added to metadata model OMV.

OntologyEvaluation Class: this class represents the evaluation of an ontology. The properties of this class are:

- evaluationComments: free text that can be used to comment the evaluation criteria;
- evaluationDate: date when the evaluation was made;
- evaluationValue: value assigned to evaluation criteria.

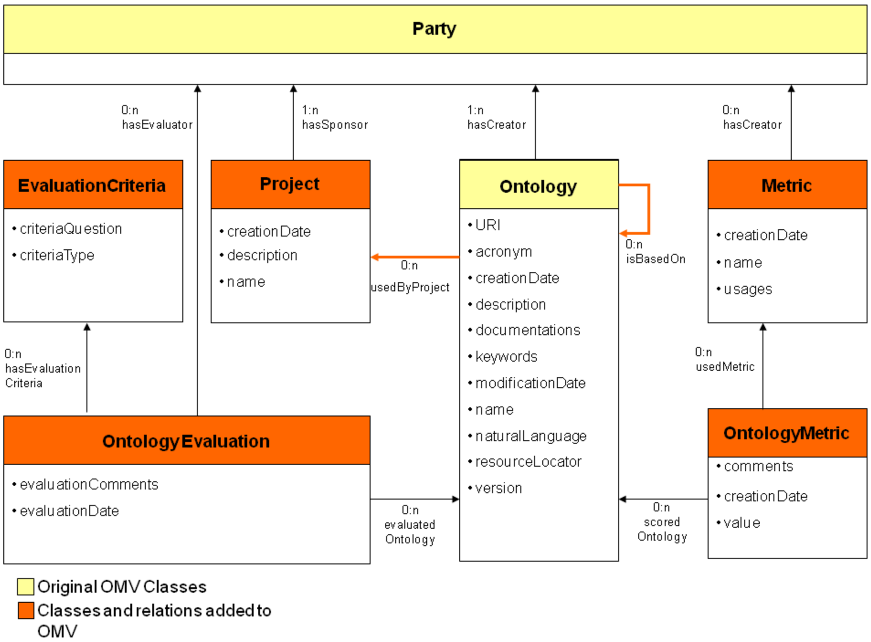

Fig. 1. Extended OMV metadata model .

EvaluationCriteria Class: this class represents the evaluation criteria used to evaluate the ontology. In our implemented model, this class has an evaluation criteria set as discussed in [16]. The properties of this class are:

– criteriaQuestion: question describing the evaluation criteria;
– criteriaType: define whether a question accept a textual or numeric answer.

Metric Class: this class represents metrics. Such metrics can be used to score an ontology, assigning a quantitative value, in order to assist users to select ontologies. In OMV-R there are 20 implemented metrics, which will be discussed in the next subsection. The properties of this class are:

– name: metric name;
– usages: free text to describe and guide the user how to use the metric as well how the metric results may be interpreted.

OntologyMetric Class: this class represents instances of the calculated metrics of an ontology. One or more metrics may be assigned to an ontology, they can help users to select ontologies on the basis of different points of view. The properties of this class are:

– comments: free text to give more information about a metric;
– creationDate: date when the metric was calculated;
– value: value assigned to a metric.

Project Class: this class represents research or commercial projects that use a given ontology or a set of them. The properties of this class are:

- creationDate: date of the project's beginning;
- description: free text to describe the project's objectives and artifacts produced by them;
- name: the project's name.

Relations: here we present the relations among these new classes.

- evaluatedOntology: relation between ontologies with their respective ratings;
- isBasedOn: reference to another ontology which was used in its construction, this reference allows to identify the reuse level that one ontology has, based on the quantity of references;
- hasEvaluator: reference to a research group or enterprise responsible by ontology evaluation;
- evaluationCriteria: reference to an evaluation criteria associated to an ontology;
- hasSponsor: reference to the institution sponsoring the project;
- usedByProject: reference to the project that uses an ontology;
- hasCreator: reference to the research group or institution that has developed the metric;
- usedMetric: reference to the metric assigned to a given ontology.

From the adaptation made, the OMV-R supports now the ontology evaluation description according to the methodologies proposed by [10], [16] and [13]. Another contribution is to allow OMV-R to store new metrics whenever is needed. These metrics may be used to evaluate an ontology over different points of view. Finally, it is expected that the extended model OMV-R provides a greater flexibility in the search of ontologies, as it offers to users additional information about reuse.

In the next section we described the set of implemented metrics.

3 Implemented Ontology Metrics

The model has been extended to accommodate a more elaborated description through metrics. In our application of the model we have implemented ways of calculating a set of metrics to describe the ontologies in the repository. The choice of metrics were based on a detailed systematic review of the literature of ontology metrics. Following the systematic review their presentation was standardized in a common representation formalism, where C is used for classes, CS is used for set classes with subclasses, CR is used for set of root classes, CL is used for set of leaf classes, $H^C(CR_j, CL_i)$ for hierarchy of classes where CR_j is a subclass of CL_i, *prop* for the function that relates the classes in a non-hierarchy and *att* for the function that relates classes with literal values (*string*), I for instances and $inst(C)$ for the function of instantiated classes [13] [12].

Besides, we use $Count$ for count, $SuperClass$ for superclass function, $SubClass$ for subclass function, Max for maximum ($MaxDepth$ and $MaxWidth$), $Depth$ for depth, $Width$ for width, Avg for average ($AvgDepth$, $AvgMaxAvg$, $AvgWidth$, $AvgAttClass$, $AvgPropClass$, $AvgPClass$, $AvgCRC$, $AvgCRCL$, $AvgCIC$, $AvgIC$) and $StdDev$ for standard deviation ($StdDevIC$).

All chosen metrics presented result in numeric values and their calculation could be implemented. The metrics were separated in three groups as follows:

1. Group 1 is simply the count of the basic elements of an ontology (classes, properties, and instances), metrics 1 to 7.

   Number of root classes [15] [1]

   $$Count(CR) = |CR| \tag{1}$$

   where: $|CR|$ is the set cardinality

   Number of leaf classes [15] [1] [14]

   $$Count(CL) = |CL| \tag{2}$$

   where: $|CL|$ is the set cardinality

   Number of classes [7] [1] [14]

   $$Count(C) = |C| \tag{3}$$

   where: $|C|$ is the set cardinality

   Number of properties (attributes) [1]

   $$Count(att) = |att| \tag{4}$$

   where: $|att|$ is the set cardinality

   Number of properties (relations) [1]

   $$Count(prop) = |prop| \tag{5}$$

   where: $|prop|$ is the set cardinality

   Number of properties [7] [1] [14]

   $$Count(att) + Count(prop) \tag{6}$$

   Number of instances

   $$Count(I) = |I| \tag{7}$$

   where: $|I|$ is the set cardinality

2. Group 2 represents the depth and width of the structure, metrics 8 to 12.

Maximum depth [7] [1] [14]

$$MaxDepth(H^C) = \text{the major ontology depth counted from each root class } (CR_j) \tag{8}$$

Average depth [1] [4] [15]

$$AvgDepth(H^C) = \text{the average ontology depth added from all root classes } (CR) \text{ by counted from each leaf class } (CL_i) \tag{9}$$

Average maximum depth to average depth [14]

$$AvgMaxAvg(H^C) = MaxDepth(H^C)/AvgDepth(H^C) \tag{10}$$

Maximum width [1]

$$MaxWidth(H^C) = \text{the major ontology width counted from each class with subclass } (CS_k) \tag{11}$$

Average width [1]

$$AvgWidth(H^C) = \text{the average ontology width added from all classes with subclasses } (CS) \text{ by counted from each class with subclass } (CS_k) \tag{12}$$

3. Group 3 represents the calculation of simple averages and standard deviation from combinations of the basic elements, metrics 13 to 20.

   Average number of properties (attributes) to total number of classes [1]

$$AvgAttClass = Count(att)/Count(C) \tag{13}$$

   Average number of properties (relations) to total number of classes [1] [14]

$$AvgPropClass = Count(prop)/Count(C) \tag{14}$$

   Average number of properties to total number of classes [1]

$$AvgPClass = (Count(att) + Count(prop))/Count(C) \quad (15)$$

Average number of leaf classes to total number of classes [1] [14] [13]

$$AvgCRC = Count(CR)/Count(C) \quad (16)$$

Average number of leaf classes to number of classes root [14]

$$AvgCRCL = Count(CR)/Count(CL) \quad (17)$$

Average number of classes populated to total number of classes [13]

$$AvgCIC(H^C) = Count(inst(C))/Count(C) \quad (18)$$

Average of the total number of instances to total number of classes

$$AvgIC(H^C) = Count(I)/Count(C) \quad (19)$$

Standard deviation of the total number of instances to total number of classes [13]

$$StdDevIC(H^C) = \sqrt{\sum (Count(inst(C)) - AvgIC(H^C))^2 / Count(I)} \quad (20)$$

The $AvgIC(H^C)$ and $Count(I)$ were calculated and used in $StdDevIC(H^C)$. It is not directly referenced in papers obtained in the systematic review.

Note that OMV documentation refer to four metrics only: number of class, number of property, number of individuals and number of axioms.

4 Application evaluation

OntoLP Portal[1] offers ontologies and ontology research related resources, with special preference to ontologies written in Portuguese language. At *Home* link the objectives of the portal are shown, whilst the link *Resources* list ontologies and related works. The link *About* contains information referring to the research group involved in this project. Through the *Contact* link others research groups in the field can send their works and suggestions. Finally, the link *Links* shows researches groups that have been colaborated directly or indirectly with the OntoLP Portal as well as events related with ontology research.

An ontology search was performed just to build the repository. We conducted searches in Google [2] specifying the type of query (e.g. "field filetype: owl"), in Swoogle [3], in OntoSelect [4] and pages of projects and research groups working with the subject. Also we announce the OntoLP portal in several Brazilian

[1] http://www.inf.pucrs.br/˜ ontolp/index.php
[2] http://www.google.com.br/
[3] http://swoogle . umbc.edu /
[4] http://olp.dfki.de/ontoselect/

research lists and organized an event with the intention of receiving these resources.

As a result we find resources in different domains such as ecology, nanotechnology, art, curriculum, emotion, privacy, network of scientific knowledge, smartphones, music and stimulus equivalence. Today the OntoLP portal has 25 ontologies. Note that the currently portal does not use any advanced structure, such as a metamodel, to store the descriptions of ontologies. Such descriptions were manually built and hard coded in HTML, which are stored in the file system on the Web server. In order to improve the current services of OntoLP portal, a new prototype application was developed that uses the OMV-R to keep the descriptions of ontologies. The next subsection will take a look deeper how this new version of OntoLP portal was developed.

4.1 Infrastructure

A tool has been developed in order to automate the receiving process. This tool reads the ontology and includes information about ontology metrics. It has also another interface for the inclusion of other metadata, to be used in the description of ontologies in the OntoLP Portal, that corresponds with instances of OMV-R. Ontologies are thus received through a submission form. Figure 2 shows the flow of information during the ontology reception process.

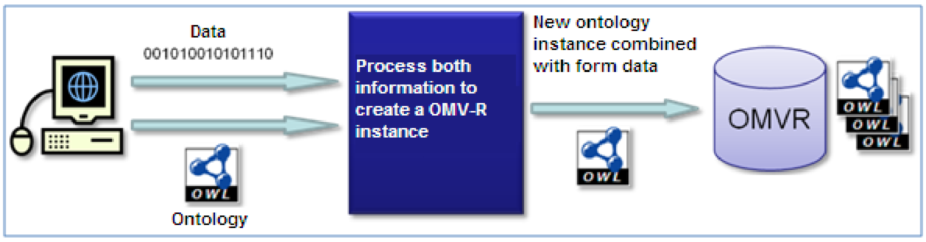

Fig. 2. Flow to store an ontology into OntoLP Portal.

Besides the ability to create OMV-R's instances, the tool offers others services such as information retrieval of the ontologies, ability to create and retrieve ontology evaluation instances as well as the metrics values for each OMV-R instance that represents a description of the ontology. The recovery of information from ontologies is made through a set of methods that read data from OMV-R and then, record the information on a XML file wich will be used by presentation layer of the prototype application. Therefore, this XML file works like a communication area between OMV-R ontology and end user interface.

Finally, the tool uses the Jena's SPARQL query engine to implement the advanced search service in the OntoLP Portal. The query is performed on the OMV-R, specifically over the Ontology class to retrieve all the ontologies that match the user's filter.

Figure 3 presents the OMV-R metric instances for the stored ontologies. Notice that the OntoLP Portal prototype application can be accessed through following link: http://www.ontolp.com.br. Both versions of portals are available for readers to discover the differences between portals (take a look at previous page on footnote for the URL of the current portal).

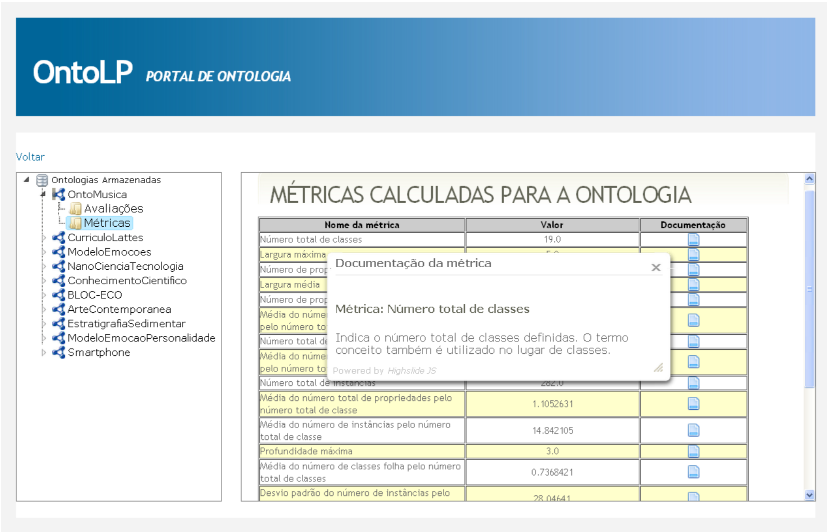

Fig. 3. Portal OntoLP interface showing an instance of OMV-R.

4.2 Evaluation

We applied two questionnaires to evaluate the adaptations made to the portal which follows the new proposed model.

The goal of the first questionnaire was to evaluate the opinion of a users' group, researchers and engineering of ontologies about the usefulness of metrics in specific repositories, in this case in OntoLP portal.

In the first phase, the questionnaire (in print format) was applied to people in national workshop about ontologies. In the final phase the questionnaire (in digital format) was applied to visitors of the OntoLP portal. The questionnaire was available for a month (December 6, 2009 until January 6, 2010).

The final result showed that 76 % respondents have developed ontologies, 83 % have searched for ontologies written in English, Portuguese, or Spanish. The language most cited was English and Portuguese was in the second position. The sample consisted of 14 PhD, 11 MSc and 5 senior experts. At last, ontologies in many domains were mentioned.

We observed that there was no preference regarding groups 1, 2 and 3, their usefulness was better appreciation as a whole. Most people consider the use of metrics important for the portal.

The respondents commented about the lack of more specific metrics (total number of relations' types, where such relations would be " is-a", " part-of") and about the lack of the presentation in percentages the results (percentage of classes that did not have superclasses). These suggestions will be considered as future work.

The goal of the second questionnaire was to get impressions about the new services of the OntoLP Portal such as description and retrieval of ontologies based on the OMV-R metadata.

Here, the sample was formed by 19 respondents. They were instructed to visit the OntoLP Portal and to evaluate the new services such as ontology submission and search. Next, the respondents were oriented to compare the services of the current version of the portal. When asked whether these new services have eased the ontology search process, about 58% of respondents agreed, 26% remained neutral and about 15% disagreed.

Despite some disagreement about how easy is to locate ontologies in the new OntoLP Portal, about 95% of respondents believe that the new ontology submission and search services are better than the original one.

In addition to the two close questions about the satisfaction and improvement about the ontology search services, the respondents were conducted to comment referring to other information that could be added to the metadata model OMV-R. About 25% of respondents have contributed with suggestions; 32 % said that the new set of the model's information was sufficient or reasonable; and 36 % didn't answer.

The suggestions are related mainly with the ontology domain, and this information is the best way to locate ontologies. Other respondent claimed that the new portal should offer information about the relationship among ontologies, as evidence: "At first glance I would say that at some point we will need a hierarchy of ontologies (when there are more than 20 or 30 ontologies) and some mechanism to relate them (meta-ontology)". Notice tha both questionaries can by accessed through the following links: http://www.inf.pucrs.br/˜ontolp/questionario.php (for metric's questions) and http://www.ontolp.com.br/questionarioOMVR.php (for metamodel and prototype application's questions)

5 Conclusion and Future Work

This paper presented an extended model based on OMV, which is called OMV-R. The new model includes key ontology metrics identified through literature review. The model was applied to OntoLP portal which aims to maintain and distribute Portuguese resources. The model is used for the description of the resources, and was also useful as a basis for new search and retrieval services. Finally, this paper discussed the surveys performed to evaluate the model, the new portal and the use of metrics, where the objective was to assess with the community the benefits and contributions of the model for ontology description and location as well as the process of reuse of such resources.

We hope to have contributed with ideas for describing ontologies that can be included in previously proposed standards. We have also shown its usefulness for the development of software that can help researchers and engineers to find and inspect available ontologies in an organized and efficient way.

Potential future work are: (*i*) adaptation of the OMV-R API to work with the OMV-R model stored in commercial database as well as free ones; (*ii*) the development of Web services for distributed applications; (*iii*) improving the advanced search interface of the new OntoLP Portal, in order to allow users query any property of the metamodel through free-text. Nowadays only three properties have been added to advanced search interface; (*iv*) building mechanisms for visualization of ontologies described in the metadata model OMV-R.

References

1. Cross, V., Pal A.: Metrics for ontologies. In: Annual Meeting of the North American Fuzzy Information Processing Society, pp. 448-453, (2005).
2. Ding, L. Finin, T. Joshi, A. Pan, R. Cost, R. S. Peng, Y. Reddivari, P. Doshi, V. Sachs, J.: Swoogle: A Semantic Web Search and Metadata Engine. In: Proceedings of the thirteenth ACM international conference on Information and knowledge management, pp. 652-659, (2004).
3. Fensel, D., van Harmelen, F., Horrocks, I., McGuinness, D., Patel-Schneider, P.: OIL: An Ontology Infrastructure for the Semantic Web. IEEE Intelligent Systems, v. 16, pp. 38-45, Mar./Apr., (2001).
4. Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J.: Modelling ontology evaluation and validation. In: The Semantic Web: Research and Applications, pp. 140-154, (2006).
5. Gruber, T. R.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human Computer Studies 43, 907-928 (1995).
6. Hartmann, J.; Bontas, E. P.; Palma, R.; Gmez Prez, A.:Demo - Design Environment for Metadata Ontologies. In: The Semantic Web: Research and Applications, pp. 427-441, (2006).
7. Lozano-Tello, A., Gmez-Prez, A.: Ontometric: A method to choose the appropriate ontology. Journal of Database Management 15, 1-18 (2004).
8. Palma, R.; Hartmann, J.; Haase, P.: OMV Report 2.4 - Ontology Metadata Vocabulary for the Semantic Web. Technical Report, p. 94, (2008).
9. Pinto,H. S, Martins J. P.: Reusing Ontologies. In: AAAI 2000 Spring Symposium Series, Workshop on Bringing Knowledge to Business Processes, pp. 77-84, (2000).
10. Simperl, E.: Reusing Ontologies on the Semantic Web: A Feasibility Study. Data and Knowledge Engineering, v. 68-10, pp. 905-925, October, (2009).
11. Smith,M. K., Welty, C. , McGuinness,D. L.: OWL Web Ontology Language Guide: W3C Recommendation 10 February 2004.
12. Tartir, S., Arpinar, I.B., Moore, M., Sheth, A. P., Aleman-Meza, B.: OntoQA: Metric-based ontology quality analysis?. In: IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, p. 9, (2005).

13. Tartir, S., Arpinar, I.B.: Ontology Evaluation and Ranking using OntoQA. In: First IEEE International Conference on Semantic Computing, pp. 185-192, (2007).
14. Yang, Z., Zhang, D., Ye, C.: Evaluation Metrics for Ontology Complexity and Evolution Analysis. In: IEEE International Conference on e-Business Engineering, pp. 162-170, (2006).
15. Yao, H., Orme, A.M., Etzkorn, L.: Cohesion metrics for ontology design and application. Journal of Computer Science 1, 107-113 (2005).
16. Yu, J. Thom, J. A. Tam, A.: Requirements-Oriented Methodology for Evaluating Ontologies. Information Systems, v. 34-8, pp. 766-791, December, (2009).