A Characterization to Aid in Ontology Reuse

Gabriel G. Nogueira, Monalessa P. Barcellos, Vítor E. Silva Souza

Ontology and Conceptual Modelling Research Group (NEMO), Computer Science Department – Federal University of Espírito Santo
Vitória, ES, Brazil

gabriel.g.nogueira@aluno.ufes.br, monalessa@inf.ufes.br, vitor.souza@ufes.br

Abstract. *Developing ontologies is not a trivial task. Reusing exiting ontologies can be helpful in this matter. However, with the high number of existing ontologies, selecting the ones that best fit the needs of the ontology engineer is still challenging. Knowing characteristics of the candidate ontologies can help the ontology engineer to better decide which one to reuse. The master research described in this paper aims to propose a set of properties to characterize ontologies in such a way that they can be easily accessed and interpreted by the ontology engineer to make decision about which ontology to reuse when developing a new ontology.*

1. Introduction

Nowadays, ontology engineers are supported by a wide range of ontology engineering methods and tools. However, building ontologies is still a complex task even for experts [Noppens and Liebig 2009]. The emergent scenario has required more comprehensive and high-quality ontologies to solve problems involving semantic issues. In this context, developing a new ontology by reusing existing ontologies may be useful. Ontology reuse allows speeding up the ontology development process, saving time and money, and promoting the application of good practices. However, ontology reuse in general is a hard research issue, and one of the most challenging and neglected areas of Ontology Engineering [Fernández-López et al. 2019]. For example, ontology engineers still face problems to find and select the right ontologies for reuse and integrate several ontologies into a new ontology [Park et al. 2011].

Fernández-López et al. (2019) point out that although ontology reuse is recommended in the community, in practice it is not yet consolidated. Factors such as language heterogeneity, deficiencies in the documentation and lack of information about the ontology are obstacles for finding and reusing ontologies. In order to properly select the ontology that best fits a certain ontology development need, it is necessary to know additional information about it. In this sense, some ontology search approaches consider structural properties based on a graph-based view of the ontology (e.g., [Alani and Brewster 2005] and [Park et al. 2011]) to help retrieve ontologies. However, it may not be easy for the ontology engineer to interpret and access such properties. There are also works that take popularity properties into account (e.g., [Ding et al. 2004]). However, a popular ontology does not necessarily indicate a good representation of the concepts it covers. Other approaches consider subjective properties that may be ambiguous and cannot be obtained automatically [D'aquin and Gangemi 2011].

Hence, it is necessary to define a set of properties that can be used to provide relevant information about ontologies and support selecting the ones more suitable for reuse when developing a new ontology. Thus, we need a set of properties that enable to characterize ontologies in such a way that aids reuse by making smoother the process of selecting ontologies for reuse and reducing the effort of the ontology engineer in this process. This problem is addressed in the research project presented in this paper. Section 2 presents a brief background, Section 3 discusses related work, Section 4 presents the research method and Section 5 presents the current state of the research and outlines future work.

2. Background

Ontology Reuse. Reusability has long been recognized as a key attribute of ontologies, yet the principles and practice of reuse remain underdeveloped. The current lack of design through reuse presents a serious problem for the ontology community. In general, reuse can be defined as the process in which available ontological knowledge is used as input to generate new ontologies [Bontas et al. 2005]. It is as a special case of design. Intuitively, it refers to the task of taking some existing ontology and manipulating it in some way in order to satisfy the design requirements. Some more specific, related, and sometimes overlapping subtypes of reuse have been defined, such as merging and alignment, integration, modular or safe reuse, and the application of ontology patterns [Katsumi and Grüninger 2016].

Currently, there are several ontologies that model the same domain (or portion of the domain), yet varying the modelling, concepts and relations between the concepts. This creates a problem in reusing existing ontologies since the ontology engineer would have to sift through several ontologies in order to select the ones to reuse [Hlomani and Stacey, 2014].

Ontology Patterns (OPs). OPs are an emerging approach that favors reuse of encoded experiences and good practices [Falbo et al. 2013]. Patterns are vehicles for encapsulating knowledge. A pattern describes a recurring problem that arises in specific contexts and presents a well-proven solution for the problem [Buschmann et al. 2007].

3. Related Work

Some works propose properties that can be used to help the ontology engineer in the process of choosing the ontology that best fits his/her necessity when developing a new ontology with reuse. For example, Buitelaar et al. (2004) propose OntoSelect, which allows for searching ontologies for a given knowledge markup task based on coverage, structure, and connectedness. Park et al. (2011) propose an approach for ontology selection and ranking based on semantic and lexical matching. It uses measures proposed in [Alani and Brewster 2005] and adds other two (relation match and taxonomy match) that can improve the selection of ontologies. Although these works measure properties to evaluate and rank ontologies, they are not devoted to support reuse. Moreover, several measures used in these works focus on the ontology structure as a graph and can be hard to be understood by less experienced ontology engineers.

Some works define properties to assess aspects of ontology quality. Although not focused on reuse, the properties can be used to evaluate an ontology and verify if it is good enough to be reused. Burton-Jones et al. (2005), for example, propose metrics (e.g., history, authority, accuracy) to assess the syntactic, semantic, pragmatic, and social aspects of ontology quality. D'aquin and Gangemi (2011), in turn, present some characteristics generally present in "beautiful ontologies". According to the authors, a beautiful ontology is one that reflects an elegant solution for modeling a problem and it is at the same time good (in terms of formal quality), usable and practicable. Have a good coverage of the targeted domain, be often easily applicable, and be structurally well designed are some of the characteristics pointed out by the authors. Since these works are concerned with ontology quality in a broader sense, they consider several properties that are difficult to be automatically obtained and do not provide an ontology repository or a search engine for ontologies.

4. Research Method

The research method adopted in this work follows the Design Science Research (DSR) paradigm, which concerns extending “human and organizational capabilities by creating new and innovative artifacts” [Hevner 2007]. It comprises the following steps [Peffers et al. 2007]: (i) *Problem identification and Motivation*, (ii) *Definition of the objectives for a solution*, (iii) *Design and Development*, (iv) *Demonstration*, (v) *Evaluation*, and (vi) *Communication*. These steps are organized in an iterative process, with three cycles: *Relevance Cycle*, *Design Cycle* and *Rigor Cycle.*

A DSR project begins with the *Relevance Cycle*, which involves defining the problem to be addressed, the requirements, and the criteria for evaluating the results [Hevner 2007], including the steps *(i)* and *(ii)*. The *Design Cycle* involves developing and evaluating artifacts or theories to solve the identified problem [Hevner 2007], comprising steps *(iii)*, *(iv)*, and *(v)*. Finally, the *Rigor Cycle* refers to using and generating knowledge [Hevner 2007], comprising the step (vi) and the use of knowledge and foundations along with the work.

5. Current State of the Research and Future Work

In this section, the current state of the work is described by following the steps of the adopted research method. At the end, we presented future work to conclude the research project.

In the *Problem identification and motivation* step, the problem was identified from the literature (e.g. [Park et al. 2011], [Alani and Brewster, 2005], and [Gangemi et al. 2005]). The problem refers to the need for selecting suitable ontologies for reuse. Thus, in the *Definition of the objectives for a solution* step, we decided to establish a set of properties to characterize ontologies to aid in ontology reuse. As requirements of the proposed set of properties, we defined that it must: (R1) be easily accessed by ontology engineers (preferably be automatically collected); (R2) be easily interpreted by ontology engineers; (R3) help ontology engineers to make decision about which ontology to reuse to meet his/her ontology development needs.

In the *Design and development* step, we aim at developing the set of properties. We started by investigating the literature and selecting ten characteristics that most should be able to be automatically obtained from the ontology OWL file. It is important to highlight that the acceptable value of the properties is determined by the ontology engineer. This initial set of properties is focused on higher level properties. Table 1 shows the current proposed set of properties.

Table 1 – Properties to characterize ontologies aiming at reuse

Name	Description	Calculation	Based on
Applicability	Indicates the effort degree (low, medium, high) necessary to use the ontology. More complex ontologies tend to require more effort to be reused.	Ontology engineers that have reused the ontology must inform (manually) the effort needed to reuse it. Applicability is obtained by assigning the values 1, 2, and 3, respectively, to the low, medium, and high effort degree and calculating the average of the values informed by different ontology engineers.	[Gangemi et al. 2005]

Table 1 – Properties to characterize ontologies aiming at reuse (cont.)

Name	Description	Calculation	Based on
Clarity	Identifies how clear and non-ambiguous is the ontology, based on the terms used to name classes and object properties. Terms that have many meanings often open more space for misinterpretation. Thus, the higher the number of different meanings of the same term, the lower the ontology clarity.	Let C the name of the class or property in the ontology. For each C, count A (the number of word senses that the term has in WordNet [Miller 1998]). Then Clarity = A/C. The return is the average number of words senses, a value next to 0 mean that the ontology has more clarity.	[Burton-Jones et al. 2005]
Consistency	Indicates the possibility of reaching contradictory conclusions in an ontology, from valid input data. When the ontology engineer creates instances of concepts of an inconsistent ontology, he/she may find invalid statement from axioms.	Obtained by using the Hermit reasoner [Shearer et al., 2008] in order to search for inconsistency inside the ontology. The reasoner returns false in case the ontology allows contradictory conclusions. Otherwise, it returns true.	[Porn et al. 2016]
Described in more than one language	Indicates if the ontology is described in more than one language and the percentage of concepts described in each used language. This property helps the ontology engineer verify if the ontology is described in the same language, he/she have used to develop the new ontology or in another language he/she is familiar with.	Obtained by verifying if the concepts' *rdfs:label* or *rdfs:comment* properties are declared using one or more language. In this case, the return is true, and the percentage of concepts written in each language is calculated.	[De Freitas et al. 2019]
Foundational Ontologies reuse	Identifies if the ontology reuses foundational ontologies and the percentage of concepts declared using classes or object properties of the foundational ontology. Reusing foundational ontologies usually indicates a well-founded ontology.	Obtained by checking if the ontology imports foundational ontologies. If so, the percentage of classes and object properties that are *rdfs:subClassOf* or *rdfs:subPropertyOf* a foundational ontology entity is calculated. Otherwise, the return is false.	[D'aquin and Gangemi 2011]
Has documentation	Indicates the percentage of classes and object properties that have comments, descriptions or labels documenting them. Well-documented ontologies are often easier to understand.	Verify if the ontology concepts contain *rdfs:label* or *rdfs:comment* explaining or presenting examples of how to use them. If so, the return is true and percentage of concepts with description or example to use is calculated. Otherwise, the return is false.	[De Freitas et al. 2019]
Has version control	Indicates if the ontology version is identified. This information allows the ontology engineer to identify the reused ontology version, even when other versions of the ontology are available.	Return true if the ontology has the *annotationProperty owl:versionInfo*. Otherwise, the return is false.	[De Freitas et al. 2019]
Has violation	Indicates if the ontology violates OWL profiles, which can increase the reasoning complexity and hamper the efficiency of a reasoner.	For each OWL Profile (DL, RL, QL, EL and Full) check if the ontology violates the profile. The return is true or false for each checked profile. Obtained by using the Hermit reasoner [Shearer et al., 2008] combined with the OWL API to check for violations.	[Obrst et al. 2007]

Table 1 – Properties to characterize ontologies aiming at reuse (cont.)

Name	Description	Calculation	Based on
Published by	Informs the person, group or organization that published the ontology. Ontologies developed by trusted people, group or organization tend to be more reliable and have some support in case of doubts.	The name of the person, group or organization is manually informed.	[De Freitas et al. 2019]
Valid IRIs	Indicates if the ontology redirect to valid IRIs. This information is useful to avoid the reuse of ontologies that have not been maintained.	Obtained by calculating the percentage of valid IRIs (i.e., IRIs that return a HTTP response that does not mean error (e.g., 200, 300)).	[De Freitas et al. 2019]

In the *Demonstration* step, we applied the initial set of properties to support OPs characterization and selection in the GoopHub tool [Reginato et al. 2019]. The tool allows ontology engineers to select goal-oriented OPs (GOOPs) for reuse (i.e., development with reuse) as well as store new OPs that are made available for future reuse (i.e., development for reuse). A GOOP consists of an ontology fragment wrapped by a goal. In other words, it refers to an ontology model fragment that can be used to achieve a goal. A GOOP can be created whether using an ontology model fragment already built (i.e., a fragment of an existing ontology can be used to achieve a goal, giving rise to a GOOP) or building the model fragment from scratch (i.e., a model fragment is built aiming to achieve a goal). In this way, ontology engineers can use the goals they want to achieve to retrieve GOOPs. Originally, the search for GOOPs in the GoopHub was based only in the goal the ontology engineer wanted to achieve. However, this type of search may result in a high number of GOOPs, requiring the ontology engineer to analyze many ontology fragments to identify the one that best fits his/her needs to develop the new ontology, which demands effort and time. Therefore, the ontology engineer may not select the best GOOP for his/her needs or may even give up the reuse. Thus, we improved the search for GOOPs by implementing a new feature that enables GOOP's characterization and extending the search to allow ontology engineers to apply filters based on the GOOPs properties, helping them in the decision of which GOOP to reuse.

Considering the results obtained in the *Demonstration* step, we will perform a new cycle of design, development and demonstration. That is, we will return to the *Design and development* step to evolve the set of properties aiming to make it more comprehensive and improve the support to ontology selection. To do so, we will continue to investigate the literature and conduct a study with ontology engineers to identify properties they consider relevant when selecting an ontology for reuse. After establishing the set of properties, we will extend the current version of GoopHub and perform a proof of concept to show the feasibility of the proposed set of properties (*Demonstration* step). After that, in the *Validation* step, the proposed set of properties will be applied in a case study to evaluation and refinement. Finally, in the *Communication* step, we will present the research results by publishing papers and the complete research project will be documented in the master thesis.

References

Alani, H., and Brewster, C. (2005). Ontology ranking based on the analysis of concept structures. In Proceedings of the third international conference on knowledge capture.

Bontas, E.P., Mochol, M., Tolksdorf, R., Case Studies on Ontology Reuse, in: Proc. IKNOW05 Int. Conf. Knowl. Manag. (Vol. 74), 2005: pp. 345–353.

Buitelaar, P., Eigner, T. and Declerck, T. (2004). OntoSelect: A dynamic ontology library with support for ontology selection. In *In Proceedings of the Demo Session at the International Semantic Web Conference*.

Burton-Jones, A., Storey, V. C., Sugumaran, V. and Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, *55*(1), 84-102.

Buschmann, F., Henney, K., Schmidt, D.C., Pattern-Oriented Software Architecture: On Patterns and Pattern Languages, Vol. 5, John Wiley & Sons, 2007.

De Freitas, M. L., Guizzardi, R. S. S. and Souza, V. E. S. (2019). GRALD: an Approach for Goal and Risk Analysis in the Development of Information Systems for the Web of Data. *J. Softw. Eng. Res. Dev.*, *7*, 4.

Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y. and Sachs, J. (2004, November). Swoogle: A semantic web search and metadata engine. In *Proc. 13th ACM Conf. on Information and Knowledge Management*.

D'Aquin, M. and Gangemi, A. (2011). Is there beauty in ontologies?. *Applied Ontology*.

Falbo, R. D. A., Guizzardi, G., Gangemi, A. and Presutti, V. (2013, October). Ontology patterns: clarifying concepts and terminology. In *Proceedings of the 4th Workshop on Ontology and Semantic Web Patterns*.

Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M. C. and Gómez-Pérez, A. (2019). Why are ontologies not reused across the same domain?. *Journal of Web Semantics*.

Gangemi, A., Catenacci, C., Ciaramita, M. and Lehmann, J. (2005, December). A theoretical framework for ontology evaluation and validation. In *SWAP* (Vol. 166, p. 16).

Hevner, A. (2007). A Three Cycle View of Design Science Research. In SJIS, v. 19.

Hlomani, H. and Stacey, D. (2014, August). An extension to the data-driven ontology evaluation. In *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*, IEEE.

Katsumi, M., Grüninger, M., What is ontology reuse?, in: Front. Artif. Intell. Appl., 2016.

Noppens, O. and Liebig, T. (2009, October). Ontology patterns and beyond: towards a universal pattern language. In *Proceedings of the 2009 International Conference on Ontology Patterns-Volume 516* (pp. 179-186).

Obrst, L., Ceusters, W., Mani, I., Ray, S. and Smith, B. (2007). The evaluation of ontologies. In *Semantic web* (pp. 139-158). Springer, Boston, MA.

Park, J., Oh, S. and Ahn, J. (2011). Ontology selection ranking model for knowledge reuse. *Expert Systems with Applications*, *38*(5), 5133-5144.

Peffers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007). A design science research methodology for information systems research. JMIS, v. 24.

Porn, A. M., Huve, C. G., Peres, L. M. and Direne, A. I. (2016). A Systematic Literature Review of OWL Ontology Evaluation. In *15th International Conference on WWW/Internet*.

Reginato, C., Salamon, J., Nogueira, G., Barcellos, M., Souza, V. and Monteiro, M. (2019). GO-FOR: A Goal-Oriented Framework for Ontology Reuse. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE