Towards Pattern-based Knowledge about Ontological Resource Reuse

Cristina Sarasua[1]

Vicomtech - Visual Interaction Communication Technologies, Paseo Mikeletegi 57, Parque Tecnológico Miramon, 20009 San Sebastián, Spain
csarasua@vicomtech.org

Abstract. Ontology reuse is still an unsolved matter on the Semantic Web. Even though ontology specification languages provide a means for knowledge reuse and ontology management technology has evolved, ontology reuse is seldom encountered in applications. The more resources there are available, the more possibilities there are for ontology reuse, and paradoxically the more difficult it is to perform ontology reuse. This thesis proposes the notion of Ontological Resource Reuse Patterns (ORRPs), which state pattern-based knowledge about ontological resources to encourage ontological knowledge comprehension and to enhance ontology selection. ORRPs provide useful knowledge to improve ontology reuse to some extent.

Key words: ontology reuse, ontoloy patterns, ontology metadata, selection

1 Problem Statement

During the last few years, the amount of available ontological resources has increased considerably and the effort undertaken by the community to build the so-called next generation Semantic Web tools and applications has led to relevant improvements in ontology engineering [1]. However, knowledge reuse, which is one of the key features of Semantic Web technologies, seems to be under exploited. Even though ontology alignment, ontology merging and ontology integration have already been studied in depth [2], there are still many open research challenges related to the three core phases of ontology reuse: discovery, selection and integration [3].

The lack of success in reuse can be due to several reasons such as the poor quality of existing ontologies, their complexity or their low modularity. Moreover, understanding and using third party ontologies can be time-consuming, especially for non-experts, and if no additional information or assistive tools are provided, developers end developing new ontologies that model similar knowledge, reinventing the wheel. Ontology documentation can enable a better understanding of ontological resources by means of human- and machine- readable descriptions, helping both in ontology selection and evaluation (which are related to each other as concluded by Sabou [4]). Recently developed systems are able to store and present annotations mainly about specification and authorship. This approach denotes an important step forward because the information

provided can be useful to filter suitable ontologies at first glance, but it does not suggest how ontologies can be (re)used. The provision of rich metadata describing not only technical details about the implementation, but also about the usage of the ontology can further improve ontology reuse.

2 State of the Art

This section is a brief introduction to the state of the art about ontology metadata and the latest ontology reuse technology. Besides traditional schema directories and plain repositories published several years ago, sophisticated technology is being constructed to manage large ontology repositories using collaborative authoring or even P2P environments. Bioportal [5] and Cupboard [6] are two remarkable systems for ontology repositories. The OOR (Open Ontology Repository) initiative [7] is a joint effort to develop an interoperability infrastructure to allow the execution of the whole metadata lifecycle. These systems have as their core element an ontology metadata standard, OMV (Ontology Metadata Vocabulary), proposed by Hartmann et. al [8], which specifies information about a useful set of ontology features regarding representation, authorship and ontology engineering issues. Simperl [9] presented the OntoMeta tool and PROMI, an automatic OMV-based metadata generator and a merging and integration platform respectively. The OMEGA algorithm [10] is a three step algorithm to harvest, extract and reuse OMV metadata from ontologies. The Linking Open Data (LOD) dataset cloud is composed of a great increasing amount of connected data sources describing miscellaneous domains in RDF. voiD (Vocabulary of Interlinked Datasets) [11] describes datasets on the Web of Data using RDFS, and it can be employed together with other vocabularies such as DC (Dublin Core). An extraordinary work related to reuse has been achieved in NeOn toolkit and NeOn methodology [12], and although pattern definition in ontology engineering is a relatively premature discipline, design patterns have been introduced by means of the Extreme Design (XD) plugin. According to Gangemi's et al.[13] definition, Ontology Design Patterns (OPs) are solutions to solve an ontology design problem that appeared several times. Content Patterns (CPs), already published at the ODP portal with other types of patterns, provide a modelling solution (to be imported into the new ontology) that satisfies the user requirements. These technologies show that there has been a great progress in the field. However, systems could provide more information about the way in which particular ontological resources can be related to reuse.

3 Proposed Approach

The PhD thesis proposal investigates how descriptive ontological information supplied by current systems can be improved, so as to provide users with best practices on how to proceed when reusing ontological resources. Ontological Resource Reuse Patterns (ORRPs), which are feasible solutions to solve the reuse of a particular ontological resource, will be generated, presented and used in order to encourage ontological knowledge comprehension and to enhance ontology selection on the Semantic Web. These patterns elucidate an exemplary solution based on ontology reuse methodologies, knowledge

extracted from ontology classification and application scenarios, and the specific ontological resource to be described. Specific ontological resources are analysed on the basis of different criteria, in order to obtain mainly three types of information: information about the manner in which it **should be reused** (according to its original purpose), information about the way it is being **currently reused** and information about its **possible usage**. ORRPs do not provide modelling solutions to user conceptual requirements (as CPs do), they describe the manner in which specific existing ontological resources can be reused.

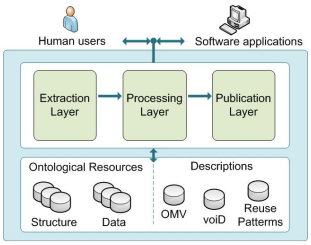

Fig. 1. The Ontological Resource Reuse Patterns Framewok

Figure 1 illustrates the concept of ORRPs. This thesis will consider ORRPs for both structure and data. Therefore, ontology repositories become as relevant as the datasets included in the LOD cloud. It is worth emphasising that ORRPs refer to ontological resources because they focus both on terminology and annotations. Nevertheless, the set of ontological resources will be limited to those that apply Semantic Web technologies. The pattern-based knowledge will offer rich descriptions, to complement existing ones, such as OMV- and voiD-based annotations on the triple space. Once the information sources are identified, the first step towards a reuse pattern will be to extract the information required for the construction of the reuse pattern (Extraction Layer). As a second step, this information will be processed in order to generate the actual reuse pattern (Processing Layer). The third step of the process consists in making the reuse pattern available on the Semantic Web (Publication Layer). Valuable knowledge for reuse will be produced considering that it could be retrieved both by human users and software applications. Consequently, ORRPs will need to be specified in a human- and a machine- readable manner. Semi-automatic techniques will be used to obtain relevant information. For example, an ontologist would browse an online repository accessing ORRP knowledge associated with multiple ontologies in order to comprehend their purpose, similarities and differences. The ontologist would understand why CIDOC CRM can help with the integration of cultural information, as well as the manner in which it could coexist with other ontologies. ORRPs could also be useful for providing a better alignment of LOD datasets for certain purposes.

Expected contributions of the PhD thesis will include a model that formally defines a reuse pattern for ontological resources, tools to generate and present reuse patterns and a simple methodology related to reuse patterns.

4 Methodology

This thesis will focus on basically three research questions: How can pattern-based knowledge be defined for describing the reuse of particular ontologies? How can this knowledge contribute to the evaluation of particular ontologies in terms of reusability? How can ontology selection processes use this knowledge? The proposed methodology is based on the iteration of six standard steps. Firstly, an analysis of the state of the art should be performed studying existing technologies and ontological resources in terms of ontology reuse, documentation and pattern definition techniques. The second step involves the definition of representative scenarios for ORRPs. Thirdly, the specification of the requirements is needed in order to identify key features that should be supported. The fourth and fifth steps entail the design, formalisation and implementation of the ORRP model, the extraction, generation and publication modules, and the interaction of the framework with existing related systems. Finally, the evaluation step will be executed to guarantee the correctness of the system and to perform qualitative and quantitative user research (in order to determine whether ORRPs are a valuable source of knowlegde for improving efficiency, quality or quantity of ontology reuse). Validation scenarios will target primarly existing multimedia ontologies, as well as ontologies related to Cultural Heritage and Tourism domains.

5 Results

Currently the objective of the thesis has been identified and the analysis of the state of the art has begun. A set of scenarios have been defined, involving different use cases where both human users and software applications can benefit from ORRPs. Since such scenarios represent the basis for the specification of an initial list of requirements, they must take into account on the one hand, multiple types of ontological resources that reuse patterns describe (e.g. LOD datasets), and on the other hand, the advantages over the state of the art. Although a list of requirements has been outlined, it should be further extended. Furthermore, this preliminary approach has been discussed with a small group of experts, although it would be desirable to receive the feedback from more experts in ontology reuse of the Semantic Web community. Moreover, the fact that modular ontologies facilitate ontology reuse and that defining modular ontologies and integrating them subsequently is considered to be a good strategy [14], is also being considered. Therefore, ORRPs should consider both ontologies as a whole and ontologies as a composition of modules.

6 Conclusions and Future Work

Ontology reuse is crucial to reach the Semantic Web. Despite the fact that current technology enables ontology reuse, this feature has not been clearly exploited. Hence, new techniques should be discovered to encourage reuse. This paper briefly presents the motivation and the proposed approach of the thesis, whose main goal is to explore how pattern-based knowledge about ontological resources can improve ontology comprehension and ontology selection on the Semantic Web. ORRPs are envisioned to be a

rich source of knowledge for ontology reuse, that together with other initiatives can achieve some progress in the field.
Future work will focus on the clarification and refinement of these initial requirements obtained. Thus, a systematic case study analysis will be carried out in order to find out some patterns within the current usage of outstanding ontological resources. Additionally, surveys among ontology experts and non-experts will be performed in order to obtain valuable information about current problems and research challenges.

References

1. Motta, E., Sabou, M.: Next generation semantic web applications. In: The Semantic Web - ASWC 2006, First Asian Semantic Web Conference, Beijing, China, September 3-7, 2006, Proceedings. (2006) 24–29
2. Kalfoglou, Y., Schorlemmer, W.: Ontology mapping: The state of the art. In: Semantic Interoperability and Integration. (2005)
3. Simperl, E.: Reusing ontologies on the semantic web: A feasibility study. Data Knowl. Eng. **68**(10) (2009) 905–925
4. Sabou, M., Lopez, V., Motta, E., Uren, V.: Ontology selection: Ontology evaluation on the real semantic web. In: Proceedings The 4th International EON Workshop, Evaluation of Ontologies for the Web, colocated with WWW2006. (2006)
5. Noy, N., Shah, N., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Montegut, M., Rubin, D., Youn, C., Musen, M.: Bioportal: A web repository for biomedical ontologies and data resources. In: The Semantic Web - ISWC 2008, Proceedings of the Seventh International Semantic Web Conference, ISWC 2008. (2008)
6. d'Aquin, M., Lewen, H.: Cupboard - a place to expose your ontologies to applications and the community. In: The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009. (2009) 913–918
7. Baclawski, K., Schneider, T.: The open ontology repository initiative: Requirements and research challenges. In: The Semantic Web - ISWC 2009 8th International Semantic Web Conference, ISWC 2009. (2009)
8. Hartmann, J., Palma, R., Sure, Y., Haase, P., Suarez-Figueroa, M.: Omv - ontology metadata vocabulary. In: Proceedings of the International Workshop on Ontology Patterns for the Semantic Web, located at the conference International Semantic Web Conference, ISWC 2005. (2005)
9. Paslaru Bontas, E.: A Contextual Approach to Ontology Reuse. PhD thesis, Freie Universität Berlin, Germany (2007)
10. Ungrangsi, R., Paslaru Bontas Simperl, E.: Omega: An automatic ontology metadata generation algorithm. In: Knowledge Engineering: Practice and Patterns, 16th International Conference, EKAW 2008. (2008) 239–254
11. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing linked datasets. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009. (2009)
12. d'Aquin, M., Gangemi, A., Motta, E., Dzbor, M., Haase, P., Erdmann, M.: Neon tool support for building ontologies by reuse. demo. In: International Conference on Biomedical Ontology, ICBO 2009. (2009)
13. Gangemi, A., Presutti, V.: Ontology Design Patterns. In: Handbook of Ontologies (2nd edition). Springer: Berlin (2009)
14. Stuckenschmidt, H., Parent, C., Spaccapietra, S., eds.: Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization. Volume 5445 of Lecture Notes in Computer Science. Springer, Berlin (2009)