



Vol-596

urn:nbn:de:0074-596-3

Copyright © 2010 for the individual papers by the papers' authors. Copying permitted only for private and academic purposes. This volume is published and copyrighted by its editors.

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

Ontology Repository for User Interaction

Martins Zviedris¹

¹ Institute of Mathematics and Computer Science, University of Latvia, Raina bulv. 29,
Riga LV-1459, Latvia
Martins.Zviedris@Lumii.lv

Abstract. The systematization and the interaction with ontologies are problems that deserve more attention. One of the key aspects is that ontology repositories should also work as the first step in a data interaction process for end-users, not only as a collection of ontology schemas. We propose a novel systematization of similar domain ontologies described by a high-level abstraction domain ontology that could be used as a domain ontology repository and access point to gather instance data.

Keywords: Ontology Systematization, Ontology Interaction

1 Introduction

Data systematization, representation and accessibility are key factors for the data usage. In the Semantic Web, the state of art for the data representation is ontology and currently ontology repositories are used to store collections of ontology schemas. Main disadvantage is that ontology repositories only allow to access ontology schema. Thus, this leads to an effect that ontologies are developed in an isolated way, as there is not provided access to the real data that would motivate to interlink them.

First, we propose that an ontology repository should contain link to instance data. Second we propose that it would be better to organize a repository around specific domain and group domain ontologies by additional domain specific meta-information. Added meta-information should be organized by a high-level abstraction domain ontology, described in more detailed in section 3. Thus, it would be easier to find similar ontologies and define ways to merge them together. Third, domain repositories would also work as a first step in data interaction for a domain expert user or an intelligent agent. The user would use the repository to select ontologies that contain the data of interest and use selected ontologies to construct relevant data queries. The repository thus becomes a bridge towards the real data.

2 Practical Experience and Proposed Solution

In practice we have encountered a problem, where we had to develop an ontology for medical researchers that describes data from different disease registries [1]. Since

instance data was originally stored in SQL databases then we had to work with SQL-like ontologies; we will present simple ontology examples below. The role of ontologies was to integrate eleven disease registries and to allow medical researchers to use ontologies as access point to the real data. Thus there was no need for elaborated ontology mechanisms. The only goal was to use ontologies to enable medical researchers to sort through vast amounts of data from registries without programmer's assistance.

A naïve approach is to merge all the registries into single, large ontology that could be stored in the data access point. We have successfully implemented the naïve approach. However, in this case medical researchers hardly comprehend the resulting ultra-complex ontology. The naïve approach needs improvement, as understandability because of ontology is crucial for medical researchers to use the ontology for actual data selection.

Based on the medical domain we will describe a more elaborate solution. The solution involves interaction of two steps. In the first step, a medical researcher has to choose registries that interest him. In the second step, he has to select and obtain data from selected registries. As each registry can also be perceived as an ontology then we need to develop a disease domain repository from where the medical researcher selects ontologies that interests him. So, in the first step the medical researcher selects registries (ontologies) from high-level abstraction disease ontology that are merged into single ontology. In the second step, the end-user can interact with data better as ontology consists of a smaller number of classes and most of them are of end-user's interest. The main idea in this approach is to develop a high abstraction level ontology that represents features from all registries. As a result, the end-user in the first step can select registries (ontologies) more conveniently and can further work with single ontology containing the needed data.

3 A Detailed Example of Medical Domain

We start with a requirement to integrate different medical disease registries into single integrated registry. Each registry contains patient's data. We propose that all registries should be stored in a medical domain repository. Also, it is required that ontology records contain links to instance data. As for medical researchers it is often necessary to work with several ontologies at a time then we also propose that these ontologies need to be organized in a more elaborate way using a high-level abstraction medical disease registries ontology.

To better understand a high-level abstraction ontology, we will build it from medical domain examples. We need to integrate two simplified registries – diabetes registry depicted in Fig 1. and cancer registry depicted in Fig 2. In practice these registries contained also other information connected to simplified solution classes and consisted of about 10 classes and about 20 enumerated classes used for classification.

By analyzing depicted registries we can identify similar structure in them. Each registry contains such general concepts as person, disease information, disease details and disease cure. A question arises whether these similarities can be used to develop a

high-level abstraction ontology. As we need to develop a high-level ontology that is used in a repository for ontology selection then we need to consider only those concepts that can ease selection process. As concept “person” does not contain information useful for ontology selection it should not be considered for a high-level ontology. Still, it could be useful to mark person concept as a concept that can be used to merge registries, thus, ontologies merging could be done at least semi-automatic.

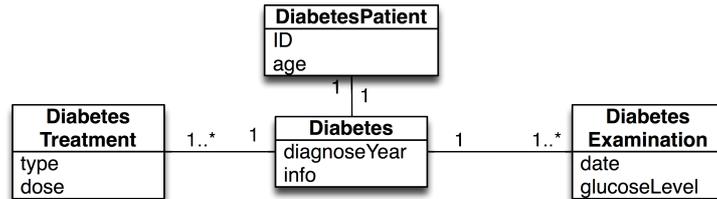


Fig. 1 Simplified diabetes registry

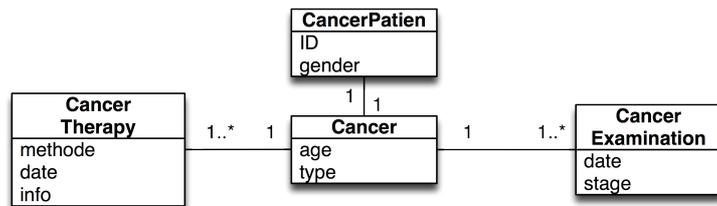


Fig. 2 Simplified cancer registry

We can identify a pattern that can be described as a disease has a treatment and an examination. The pattern is general and can be consider as a high-level abstraction ontology. The pattern can be depicted as ontology in Fig. 3. (additional information about the treatment and the examination may be added).

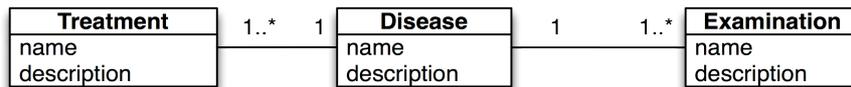


Fig. 3 Common disease ontology

Such pattern was discovered in eleven registry ontologies that were developed for medical registries. We can see that the pattern ontology is very simple and easy to grasp for the end-user. In addition, most medical disease ontologies can be described as an instance of this ontology. Still, this pattern ontology lacks meta information about actual registries and where one can find instance data and thus link to it. Also, it could be possible that data is stored in more than one place; for example, each clinic could have own cancer registry and specific disease details. We add links between the high-level abstraction ontology instances and corresponding ontologies. Links are also added towards real data and contain information about, for example, how long

data is gathered, thus links allows additional selection possibilities. Schematically structure is depicted in Fig 4.

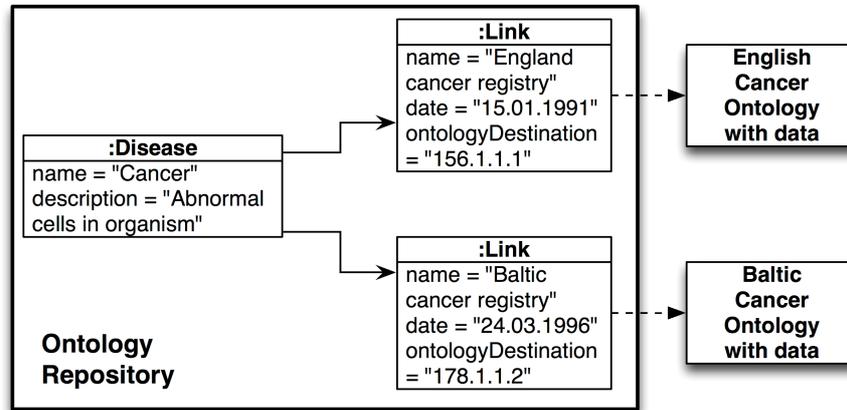


Fig. 4 Connection between ontology levels

We will sketch how a medical researcher could gather relevant data about cancer and its treatment possibilities using the proposed solution. We will not describe a specific way to query ontology as it can be done through SPARQL or more preferably by a graphical query language [2, 3]. Firstly, a medical researcher connects to a disease domain repository and queries high-level ontology depicted in Fig. 3. He restricts that he is interested in disease with name = “Cancer” and all corresponding treatments. As a result he gets instance pairs of cancer and corresponding cancer treatment and link information to ontologies that contain instance data. At this point the medical researcher can further restrict registries that interest him, for example, he could be interested only in registries that gather data at least 10 years. If he selects at least two registries, for example, the Baltic cancer registry and the England cancer registry then ontologies from both registries are merged into single ontology. This can be achieved using information that both ontologies contain similar concepts “person” and that in both ontologies is present abstract Cancer class as super class of specific Cancer classes that contain data. This ontology is presented to a medical researcher, where he can gathers clinical data for further analysis.

4 Related Work

From technical point of view, OWL2.0 allow to use punning, where an object can be represented as a class. Still, metadata addition to ontology does not solve the problem of how to group similar domain ontologies together for further interaction. Most existing ontology repositories just collect ontologies and allow to reuse ontologies. For example [4] does not give possibility for further interaction with collected ontology data that is needed for common users. They even do not collect links to existing data.

5 Results and Future Work

In practice, we have designed and implemented the naïve approach with six different disease registries [1] integrated into single ontology. To allow medical researchers query data, we have developed and implemented a graphical query language [2]. After implementing the prototype we gathered user feedback to evaluate our work. Most valuable feedback that we got was that ontology was too ultra-complex for medical researchers. Also, it was relatively easy to produce queries in ontology part, with which a researcher was familiar with. Other valuable feedback was that medical researchers are interested in registries meta information, for example, how long data has been gathered.

It would be important to practically implement and test specific domain repository with access to ontologies that contain real data. The proposed approach needs to be developed in more details to interlink between an ontology repository and ontologies that contain real data. Interesting problem also would be to find whether there should be predefined ontologies for specific diseases that could be configurable for each registry that contains such disease data and such ontology could be used for mash-up. As we have only developed the theoretical approach for medical domain it is important to go further into other domains and see possibility of a high-level abstraction ontology approach. We should mention that such domain repositories would be useful for intelligent agents as they could find links to similar data in one place.

Acknowledgements

I would like to thank prof. Guntis Barzdins for valuable discussions and prof. Karlis Podnieks for useful ideas. Also I would like to thank Arturs Sprogis and Renars Liepins for valuable assistance.

References

1. Barzdins G., Liepins E., Veilande M., Zviedris M., Ontology Enabled Graphical Database Query Tool for End-Users, Selected papers from DB&IS'2008, Hele-Mai Haav (Eds.), Frontiers in Artificial Intelligence and Applications series, IOS Press, 2009. 187:105--116
2. Barzdins, G., Rikacovs, R., Zviedris, M.: Graphical Query Language as SPARQL Frontend. In Grundspenkis, J., Kirikova, et. al. (Eds.), Local Proceedings of 13th East-European Conference (ADBIS 2009), pp. 93--107. Riga Technical University, Riga. (2009)
3. Chen, H., Wang, Y., Wang, H., Mao, Y., Tang, J., Zhou, C., Yin, A., Wu, Z.: Towards a semantic web of relational databases: A practical semantic toolkit and an in-use case from traditional chinese medicine. In Cruz, I.F., et.al, eds.: 5th International Semantic Web Conference. LNCS, vol. 4273, pp. 750--763. Springer (2006)
4. N. F. Noy, N. H. Shah, P. L. Whetzel, et. al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research, 2009