Towards a Unified Metadata Model for Semantic and Data Mappings

Sarah Alzahrani[1,2], Declan Osullivan[3]

[1]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
[2]College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia
[3]ADAPT Centre for Digital Content, School of Computer Science and Statistics, Trinity College Dublin, Dublin , Ireland

Abstract

Declarative data mappings are a valuable method for specifying how two datasets relate to one another. Creating these mappings is not a simple process, since they are typically intended for a particular data integration purpose and require expertise, resources, and time to accurately compose. However, there are still many challenges facing researchers regarding finding, accessing, understanding, and reusing these mappings, including the absence of detailed documentation and metadata related to their development lifecycle. In this paper, we argue that a unified metadata model that captures every aspect of the mapping development lifecycle is the most effective means of overcoming the current limitations

Keywords

Mappings, Alignments, Metadata Model, Knowledge Representation

1. Introduction

Within the linked data and semantic web domain mapping processes can be classified into three categories, namely interlinking mapping, uplift/downlift mapping, and ontologies mapping (sometimes called ontology alignment). The process of creating these mappings can be a complex and time-consuming task [1]. Thus, the reuse of previously developed mappings is considered necessary. It is our intuition that annotating mappings with a unified machine-processable metadata model describing the development lifecycle may enhance the discovery, reuse, and quality assessment of such mappings. This paper presents an overview of our proposed mapping lifecycle and a description of our initial metadata model.

2. Proposed Mapping lifecycle

A mapping lifecycle breaks down the mapping process into stages. The proposed lifecycle in this research has been designed such that it will be easy to follow by engineers who may not have a knowledge engineering background but who may have a more general software engineering background. The lifecycle is composed of six phases: Analysis, Mapping Design, Mapping

OM-2022: Proceedings of the 17th International Workshop on Ontology Matching, October 2022, Hangzhou, China (Virtual)

salzahra@tcd.ie (S. Alzahrani); declan.osullivan@tcd.ie (D. Osullivan)
0000-0003-2169-7006 (S. Alzahrani); 0000-0003-1090-3548 (D. Osullivan)

Table 1
Proposed metadata fields

Phases	Proposed metadata fields
Analysis	Metadata about stakeholders (URI, Name, Background, Role, organization), Metadata about the purpose of the mapping (Requirements, Type of mapping, Mapping domain, Domain assumptions, technical requirement, Risks, or issues), and Metadata to describe inputs that will be mapped (URI, Name, Source, Type, Creator, Format)
Design	(Final design decisions,Design decision justification,Quality metrics to consider during the development)
Development	(URI, Name, Start/End date of the development, Tools, Mapping method, Mapping algorithm, Format)
Testing	(URI, Name, Testing type, Date/Time, Testing result)
Maintenance	(Publisher name, Publisher source, Version number, Version date/time)

Development, Mapping Testing, and Map-ping Maintenance. We argue that the lifecycle phases are appropriate irrespective of which of the three types of declarative mappings (ontology, uplift/downlift, interlinking) is being handled, but naturally, some parts are more relevant/less relevant depending on the type of mapping.

3. Initial Metadata Model

Several studies in the literature have introduced different metadata models and ways to enrich mappings with metadata [2][3]. Generally, previous studies have developed metadata models for specific types of mappings, and a few have considered comprehensive metadata fields reflecting the stages of mapping development. In previous research such as [1] and [4], we have considered the availability of comprehensive metadata fields reflecting the different stages of mapping development for ontologies mappings. In this study, as an initial step toward supporting mapping-related activities, we have developed a unified metadata model that captures the lifecycle of mapping development. A summary of the metadata fields for each phase in the proposed lifecycle is shown in the table below.

4. Future work

The next step in our research is seeking feedback from the community through a survey. The feedback gathering is still ongoing, and 18 responses have been received to date. Your participation is appreciated at https://rb.gy/kfpcrk. Next, we'll examine how the metadata will be represented and attached to mappings. Initially, we will investigate RDF* and Named graphs as possible ways to represent our metadata model. Following on, an annotation tool will be developed, and its usability and accuracy will be evaluated by engineers. Lastly, we believe this work will contribute to building a collaborative mapping platform where semantic web and linked data community members can share, update, and evaluate mappings.

5. Acknowledgments

The first author acknowledges financial support from Imam Mohammad Ibn Saud Islamic University and the Saudi government represented by the Royal Embassy of Saudi Arabia Cultural Bureau in Dublin. The second author is partially supported by the SFI ADAPT Research Centre (grant number 13/RC/2106_P2).

References

[1] C. Debruyne, B. Walshe, and D. O'Sullivan, "Towards a Project Centric Metadata Model and Lifecycle for Ontology Mapping Governance," in Proceedings of the 17th International Conference on Information Integration and Web-based Applications Services, New York, NY, USA, 2015, vol. 1–10, p. Article 50. doi: 10.1145/2837185.2837201.

[2] A. Dimou, T. De Nies, R. Verborgh, E. Mannens, P. Mechant, and R. Van de Walle,"Automated metadata generation for Linked Data generation and publishing workflows" in LDOW2016, 2016, pp.1–10.

[3] N. Matentzoglu et al., "A Simple Standard for Sharing Ontological Mappings (SSSOM)," Data-base, vol. 2022, p. baac035, Jan. 2022, doi: 10.1093/database/baac035.

[4] H. Thomas, R. Brennan, and D. O'Sullivan, "Using the OM2R Meta-Data Model for Ontology Mapping Reuse for the Ontology Alignment Challenge: A Case Study," in Proceedings of the 7th Interna-tional Conference on Ontology Matching - Volume 946, Aachen, DEU, 2012, pp. 61–72.