ApplyOnTop - Approach for semantic data integration based on Ontop

Mônica Rocha de Ázara[1], José Maria Parente de Oliveira[1]

[1]Instituto Tecnológico de Aeronaútica (ITA), São José dos Campos, SP, Brasil

Abstract

With the increase in data, knowledge management and system integration become challenging. Although the use of Virtual Knowledge Graph, via Ontop, is a promising solution, its application is complex. The article introduces ApplyOnTop, an approach for semantic integration of databases based on Ontop. This approach proposes the execution of structured activities that assist in applying Ontop. To validate the applicability of ApplyOnTop, an experiment was conducted, using the approach in a real-world database integration scenario. The results demonstrate improvements in ontology creation and linking with databases, systematizing the use of Ontop for more effective integration.

Keywords

Ontology, Semantics, Data Integration, Mapping, Database, Computing,

1. Introduction

The increasing volume of data makes knowledge about it crucial[1] [2]. Data-driven organizations better face challenges as they can identify risks, justify situations, predict trends, and take impactful actions. In organizations, there are several obstacles to interoperability, among them being the fact that legacy systems used by companies were created independently and do not share the same semantics for the terminology of their corporate models [3]. However, complex analyses often require integrating legacy systems, which can be difficult due to lack of interoperability and query complexity. The ontology-based semantic integration approach emerges as a solution, such as Virtual Knowledge Graph and Ontop [4]. Ontop enables mapping databases to ontological elements, facilitating queries and integrations. Despite Ontop's potential, its application is challenging, requiring care in ontology development and linking to data. The approach presented in this work, ApplyOnTop, aims to structure Ontop application by defining activities to minimize the complexities of semantic data integration. Objectives include developing an approach for ontology construction and linking to databases, as well as experimenting with ApplyOnTop in a real-world context to validate its applicability.

2. Works

For this work, 20 studies adhering to its theme were selected. These studies underwent an evaluation stage, which included the application of six classification criteria listed below.

1. Did the work use a relational database as a data source?
2. Did the work generate an ontology?
3. Did the work present a detailed process for ontology construction?
4. Did the work link the defined ontology to the database?
5. Did the work enable querying the database from the semantic layer defined by the ontology?
6. Did the work define a detailed process for linking the ontology to the database?

Proceedings of the Joint Ontology Workshops (JOWO) - Episode X: The Tukker Zomer of Ontology, and satellite events co-located with the 14th International Conference on Formal Ontology in Information Systems (FOIS 2024), July 15-19, 2024, Enschede, The Netherlands

monica.azara@ga.ita.br (M. R. d. Ázara); parente@ita.br (J. M. P. d. Oliveira)

Among the analyzed studies, all incorporated the use of ontologies. Two used existing ontologies, while the others developed their own. All allowed querying of data through the generated semantic layer, with most using a database. However, none detailed the process of ontology construction and its linkage to the database.

3. ApplyOnTop - Approach for Semantic Data Integration Based on Ontop

The ApplyOnTop is an approach designed to streamline database integration through the use of Ontop. It guides users through a set of structured steps aimed at making the mapping process between ontology elements and data stored in databases more accessible.

ApplyOnTop is built with a focus on domains with multiple developed information systems. Its strategy involves constructing the ontology from macro views of the individual ontologies of each system involved in the domain. Subsequently, these macro views of the systems' ontologies are merged and detailed to generate the domain ontology, which is then linked to the systems' databases.

ApplyOnTop is divided into two distinct phases, as illustrated in the Figure 1. Phase 1, called Domain Ontology Construction, encompasses activities related to obtaining domain knowledge and elaborating its representation in the form of an ontology. Phase 2, Ontology-Database Linkage, includes activities involved in mapping the ontology and linking it to the data's physical structure.

ApplyOnTop allows for evaluation, refinement, and continuous improvement of products developed in previous stages, flexibilizing their sequence. As domain knowledge progresses, there is a need to revisit and improve products, especially the ontology, initially constructed with a business perspective and later enhanced with the technical perspective of databases.

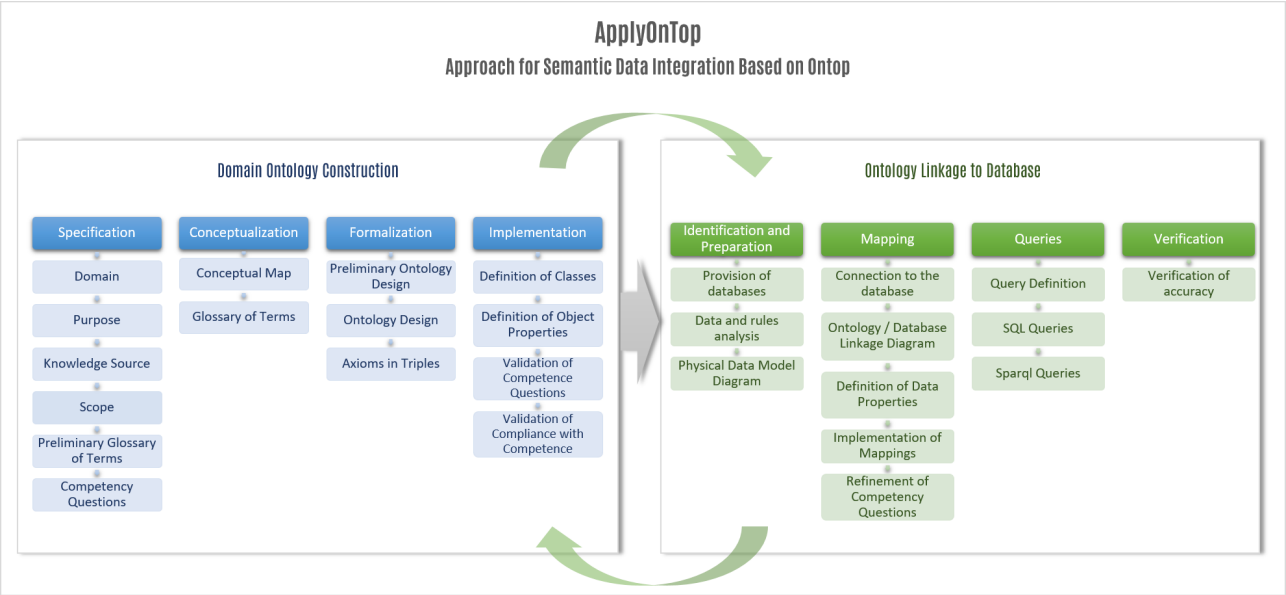

Figure 1: ApplyOnTop - Approach for Semantic Data Integration Based on Ontop

3.1. Phase 1 - Domain Ontology Construction

The ontology construction process is segmented into four distinct phases: Specification, Conceptualization, Formalization, and Implementation. In the Specification phase, the domain scope is delineated, the ontology's purpose is defined, and knowledge sources are identified [5]. In the Conceptualization phase, conceptual maps are developed for each system involved in the domain, providing a clear visual representation of the acquired knowledge. The Formalization phase focuses on transforming conceptual

maps [6] into a formal ontology, involving detailed design elaboration and axiom definition in triples. Finally, in the Implementation phase, the ontology is converted into OWL format, making it operational. This work used Protégé as the OWL editor for ontology construction [7].

The ApplyOnTop approach aims to unify the ontologies of different systems involved into a single representation, thus simplifying the integration process. After each stage, a review and refinement are conducted to ensure that the ontology meets the established requirements. This approach allows for continuous evolution of the ontology as domain knowledge deepens. At the end of the process, users have a robust and functional ontology ready to facilitate data integration and querying efficiently and effectively.

This phase was inspired by Methontology [8]. However, modifications were suggested to the original process to simplify it and adapt it to the objectives of this work. Among the changes made to Methontology, the focus on building ontologies that encompass various information systems within the same domain stands out.

3.2. Phase 2 - Ontology-Database Linkage

This phase aims to associate ontology elements with data stored in domain databases, facilitating data querying and understanding. Linkage is essential for the Ontology-Based Approach and requires creating mappings between ontology elements and tables. This phase consists of four stages: Data Identification and Preparation, Data Mapping, Query Development, and Verification.

In the Data Identification and Preparation stage, data sources are made available and analyzed. In this stage, the Physical Data Model Diagram is used to aid in understanding the data [9]. In the Data Mapping stage, the goal is to establish correspondences between ontology elements and fields in the databases[10]. This stage is carried out using Ontop[11], which allows linking the ontology to the database. To assist in mapping, the Linkage Diagram is created, representing how the ontology and database data are connected.

The Query Development stage aims to formulate queries capable of retrieving information from the databases and efficiently answering questions defined for the ontology. SQL queries compatible with the database structure and SPARQL queries to interact with the ontology and associated data are elaborated. Finally, in the Verification stage, tests are conducted to verify the accuracy of the linkage between the ontology and the databases. Throughout the process, opportunities may arise for ontology evolution to a more robust and domain-appropriate representation.

4. Experiments and Analysis of Results

For the experiment, the domain of Contracts and Customer Support was chosen. An ontology for this domain, named Ontology for Contracts and Customer Support Integration, was developed as the foundation for the experiment. Following ontology construction, steps were taken to link it to the relevant databases.

ApplyOnTop proved effective in integrating the systems within the domain, as evidenced by the results obtained throughout the proposed phases and stages. During the Domain Ontology Construction phase, a comprehensive view of the entities contained in the systems was obtained. ApplyOnTop introduced defined stages that enhanced domain understanding, initially encompassing business perspective through the use of Conceptual Map and Preliminary Ontology Design. This understanding contributed to the elaboration of the Ontology Design, covering the classes identified in the domain's systems.

An important aspect of ApplyOnTop was its flexibility and iterative format, allowing for continuous adjustments and improvements. The application of proposed diagrams in ApplyOnTop, such as Ontology Design and Ontology-Database Linkage Diagram, significantly aided in the analysis and construction of both ontology and mappings. This contribution was crucial due to the ease of understanding and visualization of the necessary relationships and linkages for enabling integration.

This outcome validates the developed approach, offering organizations an option to effectively and accurately address the complex integration demands present in their businesses. ApplyOnTop provided

a systematic view of the steps required to enable semantic integration using Ontop.

5. Conclusion

This work aimed to develop an approach for semantic database integration based on Ontop, named ApplyOnTop. This approach proposes the definition of structured activities to help users minimize the complexities inherent in semantic data integration, making the application of Ontop more systematic.

To validate the functionality of ApplyOnTop, an experiment was conducted, utilizing the need for integrating two systems within a company. The experiment was successful, demonstrating the clear creation of the ontology and its linkage to databases through the use of activities proposed in the approach.

ApplyOnTop presented significant aspects such as the creation of diagrams aiding analysis and decision-making during Ontop integration execution. For example, the Conceptual Map plays a significant role in obtaining initial domain information. Two other diagrams created were the Preliminary Ontology Design and the Ontology Design, allowing views of partial ontologies and the consolidated ontology to be built, aiding in analysis and ontology configuration in Protégé. Additionally, the construction of the Ontology-Database Linkage Diagram assists in mapping ontology elements with stored data.

Ultimately, the significant characteristic of the work was the creation of a systematic approach for Ontop usage in semantic integration. This systematization defines phases and steps to be executed, allowing for an iterative approach that enables revisiting previous steps to enhance already built products. Thus, as knowledge about the domain, business, and database is obtained, the ontology may need to evolve.

The ApplyOnTop approach presented a clear structure and defined guidelines to simplify Ontop usage. By establishing specific steps to follow, the process of ontology construction and linkage becomes more accessible and guided. Therefore, this work can assist professionals interested in working with Ontop, simplifying and facilitating the complex journey of data integration.

5.1. Contributions

1. **Detailing for Ontology Construction Using Menthontology:** Although Menthontology was adopted as the foundation, the ontology construction process has been enhanced. In this enhancement, the focus is on building ontologies in domains that encompass multiple information systems. To achieve this, steps have been included to guide the ontology builder from documenting the acquired knowledge, using conceptual mapping, to the actual construction of the ontology, supported by Preliminary Ontology Design and Ontology.
2. **Visualization Aspect of Ontology Design:** The use of Ontology Design facilitates analyzing and planning the ontology to be built in an easier way. This approach allows for a preview visualization of classes and relationships before implementation in Protégé, significantly simplifying the process. Thus, not only does it reduce the effort involved in construction, but it also creates an important input that can be used alongside the created ontology, contributing to a clearer conception.
3. **Integration of Partial Ontologies Based on Databases:** The strategy of developing macro views of the ontologies from each system involved in the domain, to later merge them and generate the final ontology, aids in the creation process. This approach helps to make the process clearer and more precise, allowing for a comprehensive understanding of the individual ontologies before their consolidation.
4. **Link Visualization:** The use of the Link Diagram enables the analysis and documentation of the mapping to be established between the ontology and the databases. This tool makes the linking process more transparent, providing a visual representation that facilitates understanding of the relationships between the ontology elements and the data in the corresponding bases.

References

[1] A. Patrizio, Idc: Expect 175 zettabytes of data worldwide by 2025, Network World 3 (2018).

[2] H. V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, C. Shahabi, Big data and its technical challenges, Communications of the ACM 57 (2014) 86–94.

[3] M. Grüninger, K. Atefi, M. S. Fox, Ontologies to support process integration in enterprise engineering, Computational & Mathematical Organization Theory 6 (2000) 381–394.

[4] G. Xiao, L. Ding, B. Cogrel, D. Calvanese, Virtual knowledge graphs: An overview of systems and use cases, Data Intelligence 1 (2019) 201–223.

[5] S. Rautenberg, J. L. Todesco, A. V. Steil, F. A. Gauthier, Uma metodologia para o desenvolvimento de ontologias, RECEN-Revista Ciências Exatas e Naturais 10 (2008) 237–262.

[6] J. D. Novak, A. J. Cañas, A teoria subjacente aos mapas conceituais e como elaborá-los e usá-los, Práxis educativa (2010) 09–29.

[7] Protégé, Protégé (2023). URL: https://protege.stanford.edu/, acessado em 24 de setembro de 2023.

[8] M. Fernández-López, A. Gómez-Pérez, N. Juristo, Methontology: from ontological art towards ontological engineering (1997).

[9] R. Elmasri, S. B. Navathe, M. G. Pinheiro, et al., Sistemas de banco de dados (2005).

[10] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, G. Xiao, Ontop: Answering sparql queries over relational databases, Semantic Web 8 (2017) 471–487.

[11] Ontop Team, OntopTutorial: Basic Setup (2023). URL: https://ontop-vkg.org/tutorial, acessado em 30 de setembro de 2023.