J. Mantas et al. (Eds.)

doi:10.3233/SHTI240680

Enhancing Healthcare Informatics: Integrating Category Theory Reasoning into OMOP-CDM Ontology Model

Rodrigo ALBARRAN[a,1] and Jean-Baptiste LAMY[a,2]
[a]*Université Sorbonne Paris Nord, LIMICS, Sorbonne Université, INSERM, UMR 1142, F-93000, Bobigny, France*
ORCiD ID: Rodrigo Albarran https://orcid.org/https://orcid.org/0009-0009-4513-4874, Jean-Baptiste Lamy https://orcid.org/https://orcid.org/0000-0002-6078-0905

Abstract. The task of managing diverse electronic health records requires the consolidation of data from different sources to facilitate clinical research and decision-making support, with the emergence of the Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) as a standard relational database schema for structuring health records from different sources. Working with ontologies is strongly associated with reasoners. Implementing them over expansive and intricate Ontologies can pose computational challenges, potentially resulting in slow performance. In this paper, we propose the implementation of a new reasoner based on categorical logic over a translation of OMOP-CDM into an ontology model. This enables enhancements to the efficiency and scalability of implementing such models.

Keywords. OMOP-CDM, Reasoning, Ontologies.

1. Introduction

The advent of electronic health records (EHR) has significantly enhanced the storage, transmission, and standardization of clinical patient data [1,2]. Yet, the multitude of EHR models and formats presents challenges for research studies and clinical decision support systems needing to interface with diverse EHR platforms.

OMOP-CDM (Observational Medical Outcomes Partnership - Common Data Model), developed by the OHDSI community, is widely recognized as an easy-to- use EHR model [3]. Demonstrated to have extensive content coverage compared to similar models [4], it has also drawn interest for potential use in clinical decision support [5].

Formal ontologies and the semantic web have gained widespread acceptance for formalizing medical knowledge, assisting formal reasoning, and encouraging data and knowledge reuse [6]. However, their implementation necessitates the use of a reasoner to ensure consistency and infer new knowledge from existing data. While certain reasoners offer robust reasoning capabilities, they also come with limitations [7]. It is

[1] Corresponding Author: Rodrigo Albarrán; E-mail: rodrigo.rodrigoalbarran@edu.univ-paris13.fr.

[2] Corresponding Author: Jean-Baptiste Lamy; E-mail: jiba@lesfleursdunormal.fr.

essential to assess whether they meet the requirements of the specific ontology and application context.

In this work we start from the work of Jean Baptiste et. al. [8], where they translate the OMOP-CDM model into an OWL. Based on this translation, the objective of this work is to extract a sub-ontology from the ontology they developed. Upon this sub-ontology, we implement and deploy a new algorithm aimed at enhancing the ontology's semantic expressiveness, inference capabilities, consistency and efficiency, ultimately leading to improved data management, analysis, and decision support. Our aim is to enhance the capabilities provided by OMOP-CDM, by expecting to bolster the model's capacity not only for data management and analysis but also for decision support.

2. Material and Methods

2.1. Translating OMOP-CDM Electronic Health Records to OWL Ontology

We use the ontology model based on the OMOP-CDM version 6.0 [3,8]. According to the translation from data model to ontology, People, events and expositions are the tables presented in OMOP-CDM, they are represented as concepts in the ontology model.

The OMOP-CDM database model facilitates the derivation of Eras from Drug Exposures and Condition Occurrences, with both Eras and lower-level entities linked to medical terminology concepts. Using a Python script, they parsed the OMOP-CDM specification to generate the OWL ontology, leveraging the Owlready ontology-oriented programming module [9-11]. Through automatic translation, each table in the OMOP-CDM model was converted into a class, with identifier fields mapped to object properties and non-identifier fields to data properties. Additionally, SQL data types were transformed into XML Schema data types, and universal and existential class restrictions were applied to mandatory fields in the OMOP-CDM model.

2.2. DL-LiteR Sub-Ontology Extraction

In order to improve the efficiency of the reasoner at the time of execution with respect to the ontology. The specific scope for subtracting the sub-ontology from the OMOP-CDM Ontology is identified. This was done by setting a criteria based on the requirements of the DL-Lite_R language in which the reasoner was developed [9]. Since this language only accepts: existential quantifier, inverse roles and negation, the latter for both concepts and roles.

Manually defined superclass categories such as Base Person, Eras, Events, and Durations were included. The term "Era" encompasses classes representing periods of time, including Drug Era, limited to a maximum of 3 months before renewal, and Condition Era, describing temporary conditions like fractures. Specific super-relations were also manually introduced into the ontology using a similar process. Additionally, an automated Python procedure was employed to transfer shared attributes among subclasses of a superclass to that superclass. For instance, if both Person and Provider subclasses have a gender attribute within the Base Person superclass, this attribute will be consolidated within the Base Person class through the procedure [8].

According to the direction specified by the OMOP model, the relationship between Person and Drug Era is established in the direction of *Drug Era → Person*, identified through the “*person id*” field. This directionality arises because each Drug Era is linked to a single Person in (*one−to−one*) relation; while a person may be associated with several drugs in (1^{-} *one−to−many*). Which allows the ontology to introduce an inverse relation, where from a given relation with one direction it is possible to extend to the ontology itself, by describing the opposite direction in terms of the same inverse relation.

Specific super-relations were manually introduced into the ontology, employing the same manual process used for adding superclasses. These super-relations capture all relationships between a class and all classes contained within a superclass. For instance, there are three relationships between the class Person and all the subclasses of the class Era. The relationships *has_Condition_Era, has_Drug_Era* and *has_Dose _Era*, are contained within the super-relationship *has Era*.

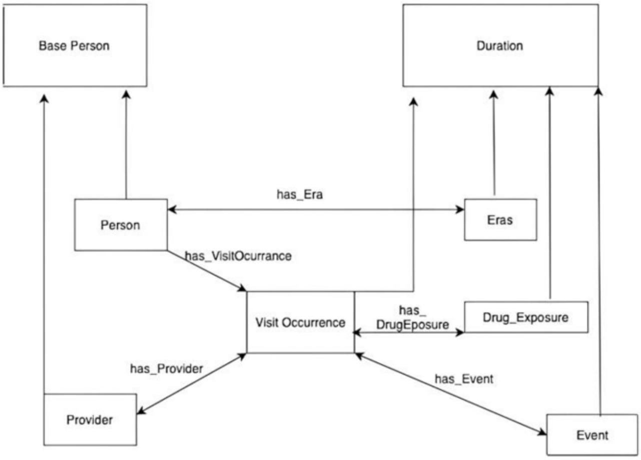

Figure 1. UML diagram showing some of the DL-Lite ontology translation of OMOP-CDM.

Figure 1 illustrates the sub-ontology model extracted from the OMOP-CDM ontology model presented in [8]. Vertical arrows depict class inheritance, while horizontal arrows denote super-relations among certain concepts, including inverse relations between some of them.

By reviewing the ontology, the classes and relationships that encapsulate the schema given by OMOP-CDM are identified. The hierarchical structure is analyzed by determining the relationships between concepts and roles. Additionally, the superclass-subclass relationships for concepts and the domain-range relationships for roles are determined.

Finally, the relevant concept and role definitions were extracted to establish the following TBox:

$$Person \sqsubseteq BasePerson \; (1)$$
$$\exists Has\,\underline{P}rovider^{-} \sqsubseteq BasePerson \; (2)$$
$$\exists Has\,\underline{E}ra^{-} \sqsubseteq Duration \; (3)$$
$$\exists Has\,\underline{D}rugExposure^{-} \sqsubseteq Duration \; (4)$$
$$\exists Has\,\underline{V}isitOcurrance^{-} \sqsubseteq BasePerson \; (5).$$

The extraction of the TBox from the OMOP-CDM schema enables the synthesis of the ontological model, utilizing both the TBox and its instances as input for the reasoning algorithm.

2.3. Reasoner implementation

In order to verify the consistency of this ontology, it is proposed to employ the IsSatisfiable algorithm, as suggested in [9]. This algorithm takes as input a $DL-Lite_R$ ontology and operates by establishing equivalence between category theory and set theory, enabling visualization of the ontology in terms of two classes. Each class is defined by the equivalence between category semantics and set semantics, along with specificity within the context of category semantics. These classes include the concept class and the role class, representing categories accordingly.

The relations between these classes represent the various hierarchies within the ontology. This process makes it possible to generate and visualize the classes and the relationships between them by means of tables. Compared to other reasoner methodologies, this algorithm assesses consistency by averaging the search for a specific element in the class hierarchy table, representing the inclusion of an empty instance. If such an element is found, it indicates ontology inconsistency.

3. Results

Figure 1 depicts the general ontology model, illustrating inheritance among superclasses and bidirectional object properties among classes. Utilizing the IsSatisfiable algorithm, applied to the sub-ontology extracted from OMOP-CDM [8,9], facilitated new reasoning processes, enabling the identification of intricate patterns and relationships. Initial tests on low-level instances formatted in OMOP-CDM revealed the consistency of the extracted sub-ontology. Subsequently, a practical test was conducted using a Python based reasoner in development. This test evaluated two scenarios: one with the complete ontology and the other with the sub-ontology, following rule reduction based on DL $Lite_R$. The performance of the reasoner with the general rules was 3.59 seconds, whereas with the sub-extracted ontology, it improved to 1.73 seconds. However, it's imperative to acknowledge that the performance enhancement observed with a small-scale ontology may not persist upon extension.

4. Discussion

OMOP-CMD was initially designed to consolidate clinical data from various EHR sources to streamline clinical research. However, it also supports clinical decision making by integrating data from different models. Using ontologies to organize EHRs benefits clinical research by providing standardized terminology, data representation, and promoting interoperability among healthcare systems, databases, and studies.

According to the guidelines established by OMOP-CDM and DL theory (on which the construction of reasoners is based), the ontology was adjusted to the most optimal DL-Lite language to verify its consistency. The test to verify the consistency of the ontology was carried out in two stages. The first stage was theoretical, using the [specific algorithm] and an example [example details]. The second stage involved a practical test using a new reasoner that is still under development, written in Python. This test considered two cases: the first with the complete ontology and the second with the sub-ontology extracted by following and reducing the rules based on DL-Lite$_R$. The results showed an improvement in the performance of the reasoner in the second case, reducing the execution time by half. It is worth mentioning that we are currently working with a small-scale ontology, which does not guarantee that the performance improvement will prevail once the ontology is extended. By improving the reasoning process, the implementation of ontology searches will allow for greater accessibility and the extraction of implicit knowledge found in ontologies.

5. Conclusion

This study initiates with the translation of the OMOP-CMD relational database into an OWL model and the extraction of a sub-ontology from this model, followed by an assessment of its consistency using a novel reasoning model. While the algorithm's current implementation remains at a theoretical level, it holds promise for enhancing analytical capabilities and offering valuable insights. The practical test demonstrated a significant improvement in the reasoner's performance when utilizing the sub-ontology, resulting in a reduction of execution time by more than half. This underscores the potential benefits of applying DL-Lite$_R$ rules to streamline reasoning processes. However, further testing with larger ontologies is imperative to validate the scalability and consistency of these performance enhancements. Ultimately, enhancing the reasoning process through ontology queries stands to offer greater accessibility and facilitate the extraction of implicit knowledge from ontologies.

References

[1] Kataria S, Ravindran V. Electronic health records: a critical appraisal of strengths and limitations. Journal of the Royal College of Physicians of Edinburgh. 2020;50(3):262-8. doi:10.4997/JRCPE.2020.309.

[2] Nordo AH, Levaux HP, Becnel LB, Galvez J, Rao P, Stem K, et al. Use of EHRs data for clinical research: historica progress and current applications. Learning health systems. 2019;3(1):e10076. doi: https://doi.org/10.1002/lrh2.10076.

[3] Reich C BRNKBC Ryan P. OMOP Common Data Model Specifications; 2018.

[4] Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. Journal of biomedical informatics. 2016;64:333-41. doi: https://doi.org/10.1016/j.jbi.2016.10.016.

[5] Gruendner J, Schwachofer T, Sippl P, Wolf N, Erpenbeck M, Gulden C, et al. KETOS: Clinical decision support and machine learning as a service–A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services. PloS one. 2019;14(10):e0223010. doi:https://doi.org/10.1371/journal.pone.0223010.

[6] Schulz S, Jansen L. Formal ontologies in biomedical knowledge representation. Yearbook of medical informatics. 2013;22(01):132-46. doi:http://dx.doi.org/10.1055/s-0038-1638845.

[7] Singh G, Bhatia S, Mutharaju R. OWL2Bench: a benchmark for OWL 2 reasoners. In: The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II 19. Springer; 2020. p. 81-96. doi:https://doi.org/10.1007/978-3-030-62466-8 6.

[8] Lamy JB, Mouazer A, Sedki K, Tsopra R. Translating the observational medical outcomes partnership common data model (OMOP-CDM) electronic health records to an OWL ontology. Stud Health Technol Inform. 2022;290:76-80. doi:10.3233/shti220035.

[9] Albarran R, Le Duc C. Reasoning in DL-L ite R Based Knowledge Base Under Category Se- ´ mantics. In: Mexican International Conference on Artificial Intelligence. Springer; 2023. p. 253-70. doi:https://doi.org/10.1007/978-3-031-47765-2 19.