Fachrepositorium Lebenswissenschaften (FRL)

Repository for Life Sciences

Temporal Data Modeling Evaluation in Knowledge Graphs: A Healthcare Use Case

Hooshafza, Sepideh | Stephens, Gaye | Little, Mark | Yaman, Beyza

Version: Postprint (Verlagsversion)/Postprint (Publisher Version)
Typ/Type: Kongressschrift/Conference Proceeding
Jahr/year: 2024
Quelle/Source: https://repository.publisso.de/resource/frl:6441256
Schlagwörter/Keywords: knowledge graph, Temporal data modeling, Evaluation framework

Zitationsvorschlag/ Suggested Citation:
Hooshafza, Sepideh et al. (2024): Temporal Data Modeling Evaluation in Knowledge Graphs: A Healthcare Use Case. International SWAT4HCLS Conference 2024. DOI: 10.4126/FRL01-006441256

Nutzungsbedingungen:
Dieses Werk ist lizensiert unter einer Creative Commons Lizenz:
https://creativecommons.org/licenses/by/4.0/

Terms of use:
This document is licensed under creative commons license:
https://creativecommons.org/licenses/by/4.0/

Temporal Data Modelling Evaluation in Knowledge Graphs: A Healthcare Use Case

Sepideh Hooshafza

ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Gaye Stephens

ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Mark A.Little

ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
Trinity Health Kidney Centre, Trinity College Dublin, Dublin, Ireland

Beyza Yaman

ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Introduction

Modelling temporal healthcare data supports medical professionals to comprehend disease patterns, evaluate patient histories, identify relationships in clinical events, and make informed predictions for improved patient care [1, 2]. A number of RDF-based approaches to modelling temporal data in KGs exist, including “standard reification”, and “singleton property”[3, 4]. However, to the best of our knowledge, the current RDF based temporal data modelling approaches have not been systematically evaluated using a healthcare use case. Existing approaches have produced different results in terms of complexity, and performance which require further evaluation and comparison [5].

Evaluation Framework

An evaluation framework is proposed to evaluate temporal data modelling approaches in KGs. The evaluation framework components consist of six phases and their associated artefacts. Each phase is designed to contribute to the evaluation of temporal data modeling approaches in KGs. Figure 1 depicts the components of the evaluation framework. The evaluation actions are displayed in white rectangles, while the artifacts are highlighted within grey shapes.

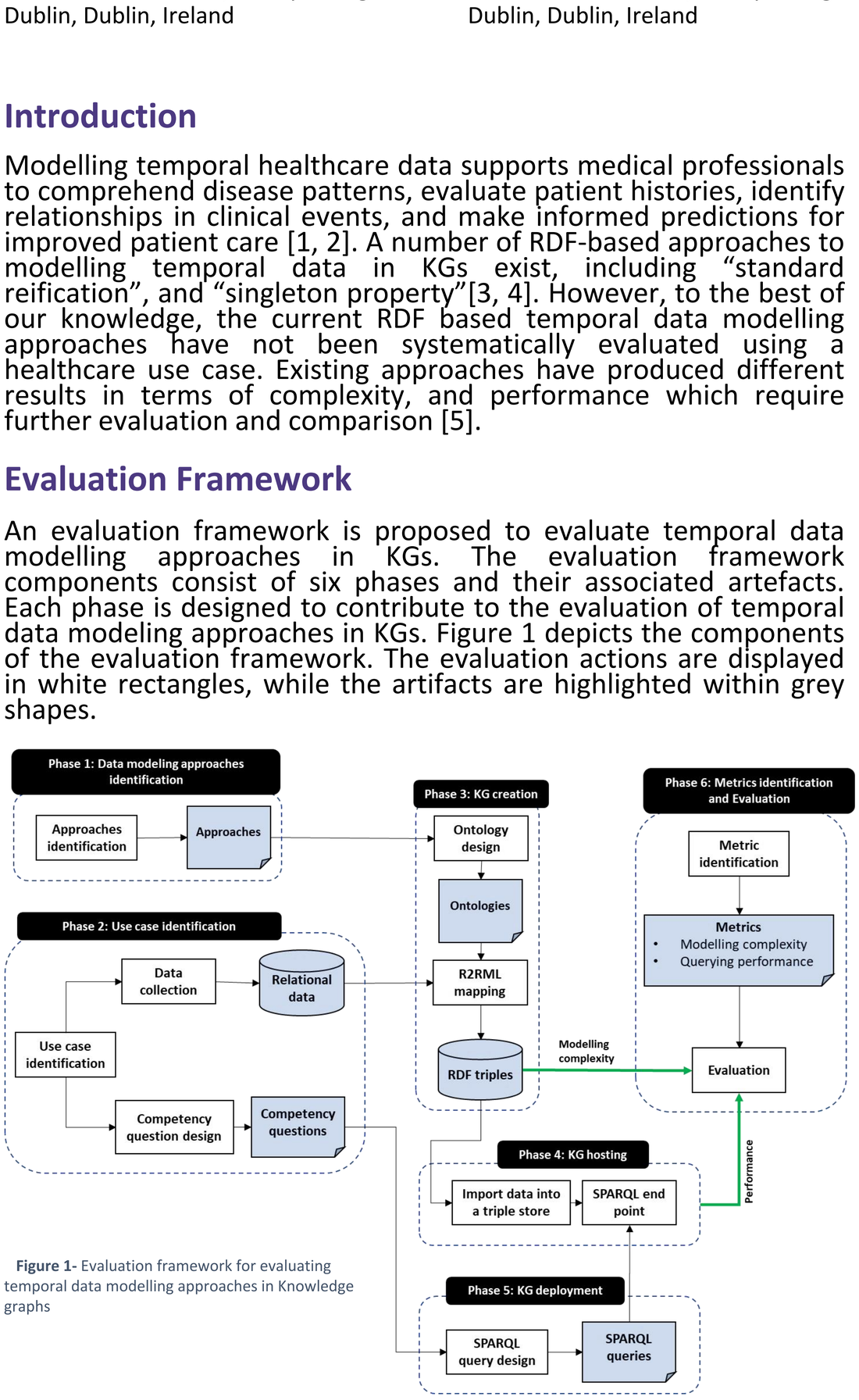

Figure 1- Evaluation framework for evaluating temporal data modelling approaches in Knowledge graphs

Case Study - FAIRVASC Project

FAIRVASC [6, 7] is a research project of the European Vasculitis Society and RITA European Reference Network . The Rare Kidney Disease Registry and Biobank (RKD) is one of the seven FAIRVASC registries and is also the source registry for this paper. The evaluation framework was applied to a healthcare use case that addresses modeling medication data for patients with the rare disease anti-neutrophil cytoplasmic antibody (ANCA) Associated Vasculitis (AAV). This study concentrated on a subset of the RKD registry. The dataset contains a total of 600 patients.

Experiment

Two well-known approaches for adding temporal data to a KG were chosen including singleton property and standard reification. The evaluation was performed based on the six phases of the evaluation framework and the healthcare use case. The data included details regarding the medications utilized for patients, including both the start and stop dates for each medication. A temporal ontology was designed based on identified approaches, RDF data was generated using the R2RML engine using the designed ontology. RDF data were imported into a triple store and queried using SPARQL.

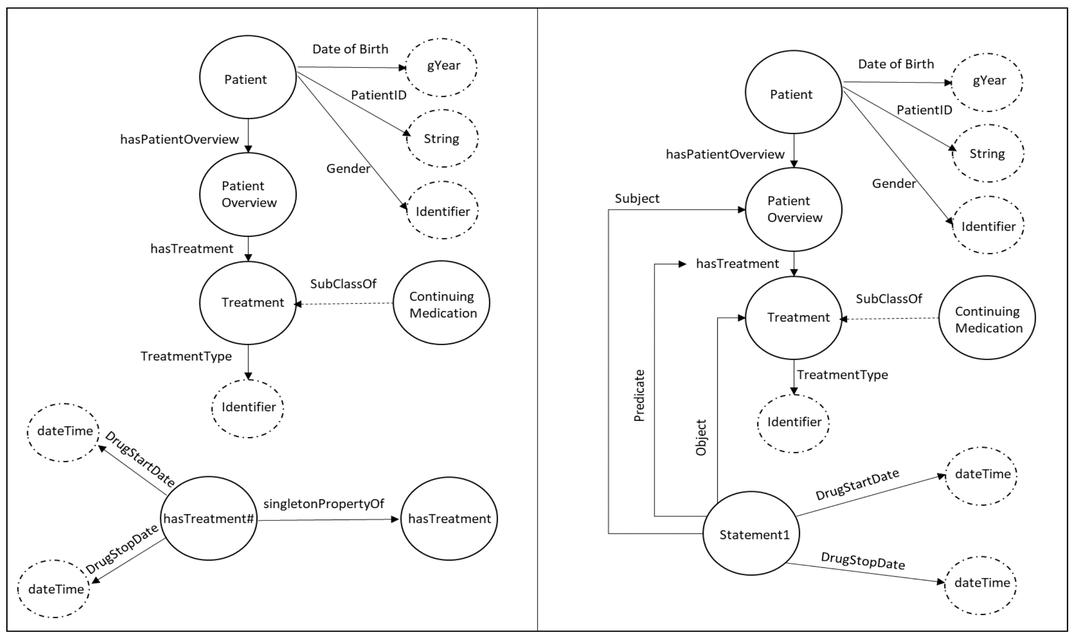

Figure 2- Ontology based on Singleton property (Left side) and standard Reification (Right side) data modelling approaches

Results

The comparative analysis between the singleton property and standard reification approaches provides insights into their performance and complexity within the realm of RDF data modeling. In terms of modeling complexity, the singleton property method exhibited a lower number of statements and generated fewer additional triples when incorporating temporal aspects. It also demonstrated reduced resource redundancy compared to the standard reification approach. While assessing performance metrics, there isn't a notable difference between the two approaches, primarily attributed to the limited size of the dataset. The assessment of singleton property and standard reification approaches provided insights into their capabilities and limitations in handling temporal healthcare data within RDF-based KGs. However, it's important to note that these experiments do not aim to provide an exhaustive and comprehensive evaluation, as this would be out of scope of this paper. To achieve thoroughness, multiple triple stores would need testing using the proposed evaluation framework and identical experimental settings.

Conclusion and Future Research

This study proposed an evaluation framework for evaluating temporal data modeling approaches in KGs. The results of this research aim to guide data and knowledge engineers in evaluating various temporal data modeling approaches within KGs. With this knowledge, they will be able to choose the methods that will best meet their needs when modeling temporal health data in graph databases. Such choices contribute to the advancement of designing graph databases for health information systems. Future research should aim to enhance the evaluation framework by encompassing a wider array of healthcare scenarios, refining the metrics used, and conducting scalability assessments with larger datasets. This will help ensure a more comprehensive understanding of how well the framework performs across various healthcare situations and when handling substantial amounts of data.

References

1. N Poh, N., S. Tirunagari, and D. Windridge. Challenges in designing an online healthcare platform for personalised patient analytics. in 2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD). 2014
2. Combi, C., G. Cucchi, and F. Pinciroli, Applying object-oriented technologies in modeling and querying temporally oriented clinical databases dealing with temporal granularity and indeterminacy. IEEE Trans Inf Technol Biomed, 1997. 1(2): p. 100-27.
3. Hernández, D., A. Hogan, and M. Krötzsch. Reifying RDF: What works well with wikidata? 2015.
4. Nguyen, V., O. Bodenreider, and A. Sheth. Don't like RDF reification? Making statements about statements using singleton property. 2014.
5. Magkanaraki, A., et al. Benchmarking RDF schemas for the Semantic Web. in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2002
6. Yaman, B., et al., Towards A Rare Disease Registry Standard: Semantic Mapping of Common Data Elements Between FAIRVASC and the European Joint Programme for Rare Disease
7. McGlinn, K., et al., FAIRVASC: A semantic web approach to rare disease registry integration. Computers in Biology and Medicine, 2022. 145.

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at TCD [13/RC/2106_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.