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

Sepideh Hooshafza[1] , Gaye Stephens[1] , Mark A. Little[1,2] and Beyza Yaman[1]

[1] ADAPT Research Centre, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
[2] Trinity Health Kidney Centre, Trinity College Dublin, Dublin, Ireland

Abstract
Healthcare data, such as patients' symptoms, laboratory test results, and various clinical measurements are temporal in nature, and are associated with a time. Modelling temporal healthcare data could benefit healthcare practitioners in healthcare decision making and support patient care. One method for modelling data that has been used in academia and industry is RDF-based knowledge graphs (KGs). Many approaches have been proposed to model temporal data in RDF-based KGs which have not been evaluated systematically in the healthcare domain. In this paper, an evaluation framework is proposed for the evaluation of temporal data modeling approaches in KGs and has been applied in a healthcare use case.

1. 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]. In this study, an evaluation framework is proposed to evaluate temporal data modelling approaches in KGs. The evaluation framework components consist of six phases including data modeling approaches identification, use case identification, KG creation, KG hosting, KG deployment, metrics identification and evaluation. 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) in FAIRVASC [6, 7] project. The framework will guide data and knowledge engineers in evaluating various temporal data modeling approaches within KGs.

2. 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 dataset contains a total of 600 patients. The data included details regarding the medications utilized for patients, including both the start and stop dates for each medication. Two ontologies were designed based on identified approaches, RDF data was generated using the R2RML engine. RDF data were imported into a triple store. A competency question was designed and RDF data was queried using SPARQL.

SWAT4HCLS 2024: The 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Science, February 26–29, 2024, Leiden, Netherlands
sepideh.hooshafza@adaptcentre.ie (S. Hooshafza); gaye.stephens@adaptcentre.ie (G. Stephens); mark.little@adaptcentre.ie (M. Little); beyza.yaman@adaptcentre.ie (B. Yaman)
0000-0002-1061-1572 (S. Hooshafza); 0000-0001-5384-6139 (G. Stephens); 0000-0001-6003-397X (M. Little); 0000-0003-2130-0312 (B. Yaman)

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 standard reification approach is more complex than the singleton property approach. While assessing performance metrics, there isn't a notable difference between the two approaches, primarily attributed to the limited size of the dataset.

Table 1- Experiments results based on modelling complexity and performance.

Category	Metric	Singleton property	Standard reification
Modelling Complexity	Number of Statements	22904	36735
	Additional triple generation	13544	27375
	Resource redundancy	21137	34479
Performance	Data load time in triple store	0.5s	2s
	Query length requirement to execute a particular task	9	9
	Query response time	6.6 s	0.2 s

3. Conclusion

This study proposed an evaluation framework for evaluating temporal data modeling approaches in KGs. The framework can 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.

Acknowledgements

This work was funded by the ADAPT Centre for Digital Content Technology under the SFI Research Centres Programme (Grant13/RC/2106-P2).

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**.