Towards A Unified Knowledge Graph Data Management System

Baozhu Liu, Xin Wang, Pengkai Liu, Sizhuo Li
College of Intelligence and Computing, Tianjin University, Tianjin, China
{liubaozhu,wangx,liupengkai,lszskye}@tju.edu.cn

ABSTRACT

Knowledge graph currently has two main data models: RDF graph and property graph. The query language on RDF graph is SPARQL, while the query language on property graph is mainly Cypher. Different data models and query languages hinder the wider application of knowledge graphs. In the paper, we propose a unified interoperable knowledge graph database system, which can effectively manage both RDF and property graphs.

Reference Format:
Baozhu Liu, Xin Wang, Pengkai Liu, Sizhuo Li. Towards A Unified Knowledge Graph Data Management System. In the 2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores (SEA Data 2021).

1 INTRODUCTION

With the proliferation of Knowledge Graphs (KG), the applications of KGs have a rapid growth in recent years. RDF (Resource Description Framework) graph and property graph are the two mainstream data models of KGs. On one hand, RDF has become the World Wide Web Consortium recommendation to represent KGs, and is widely used by triple stores, such as gStore [1]. On the other hand, property graphs are widely applied to graph databases, such as Neo4j [2]. It has been widely recognized that it is necessary to unify the data models and query languages for KG database management. To this end, we propose a unified KG data management system, which consists of three components, i.e., storage manager, query processing coordinator, and Web interface, making multiple KGs managable in a unified database management system. The queries will be translated into unified semantics denoted by relational algebra using the query processing coordinator. In storage manager, RDF graphs and property graphs will be shred into relations with the specific approaches.

2 APPROACH AND NOVELTY

As shown in Fig. 1, to the best of our knowledge, the system proposed in this paper is the first KG database system that realizes a unified storage scheme, facilitates the interoperability of SPARQL and Cypher, and meanwhile provides a Web interface to visualize the query results and explanations. (1) Based on the relational model, a unified storage scheme is utilized to efficiently store RDF graphs and property graphs, and support the query requirements of knowledge graphs. (2) Using the characteristic-set-based method,

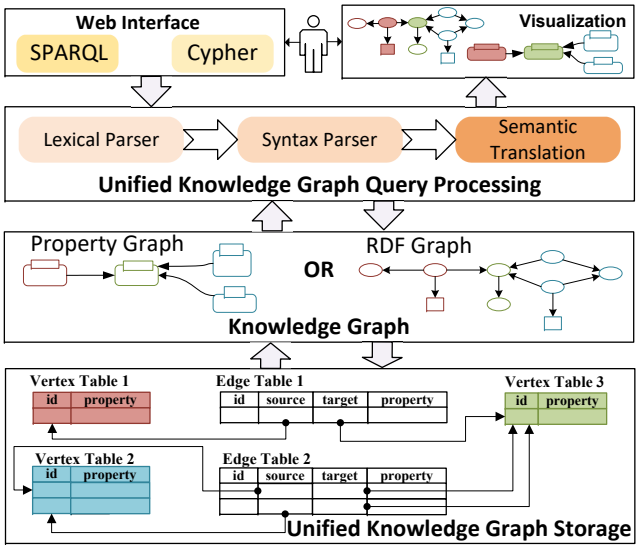

Figure 1: The Overall Architecture.

the storage problem of untyped entities is addressed. (3) The interoperability of SPARQL and Cypher is realized, and enables them to interchangeably operate on the same knowledge graph. (4) With a unified Web interface, users are allowed to query with two different languages over the same KG and visualize query results and explanations.

Due to the unified storage scheme and query processing method that we utilized, it is easier to manage multiple KGs in one database. Users no longer need to switch among different database systems to obtain storage and query support for different data models.

To verify the effectiveness and efficiency of the proposed system, extensive experiments were conducted on several data sets. The experimental results show that our system outperforms gStore [1] and Neo4j [2], which are two state-of-the-art KG database systems. The comparison of the features supported by the systems is shown in Table. 1.

Table 1: System Comparison.

System	Storage		Query			
	RDF	Property Graph	BGP	Text Search	Graph Analysis	RPQ
ours	√	√	√	√	√	√
gStore	√	×	√	×	√	×
Neo4j	×	√	√	√	√	×

3 FUTURE WORKS

In order to meet the storage and query requirements of large-scale KG data, we will focus on the distributed KG data management systems in the future. Moreover, more query features will be supported in the unified KG management system.

Copyright © 2021 for the individual papers by the papers' authors. Copyright © 2021 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).
Published in the Proceedings of the 2nd Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores, co-located with VLDB 2021 (August 16-20, 2021, Copenhagen, Denmark) on CEUR-WS.org.

CEUR-WS.org/Vol-2929/poster2.pdf

ACKNOWLEDGMENTS

This work is supported by National Key Research and Development Program of China (2019YFE0198600); the National Natural Science Foundation of China (61972275, 61972402); and CCF-Huawei Database Innovation Research Plan (CCF-Huawei DBIR2019004B).

REFERENCES

[1] Lei Zou, M Tamer Özsu, Lei Chen, Xuchuan Shen, Ruizhe Huang, and Dongyan Zhao. gstore: a graph-based sparql query engine. *The VLDB journal*, 23(4):565–590, 2014.

[2] Justin J Miller. Graph database applications and concepts with neo4j. In *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*, volume 2324, 2013.