Property graph data need a serialization

Dominik Tomaszuk1[0000–0003–1806–067X], Renzo Angles2,3[0000–0002–6740–9711],
Łukasz Szeremeta1[0000–0002–6635–0623], Karol Litman1, Diego Cisterna3
1Institute of Informatics, University of Bialystok, Ciołkowskiego 1M, 15-245
Białystok, Poland {d.tomaszuk,l.szeremeta}@uwb.edu.pl
2Department of Computer Science, Universidad de Talca, Curicó, Chile
rangles@utalca.cl
3Center for Semantic Web Research, Santiago, Chile
dcisterna@live.com

Data serialization is fundamental in data management as it allows to support database exchange, systems benchmarking, data visualization and presentation. Hence, graph serialization is critical for the evolution of graph-oriented systems and applications.

Given the popularity of database systems based on property graphs, we focus on the serialization of data conforming such data model. That is why the authors of this submission, who are members of academia, have come together to create an informal research group. Our statement is that

Property Graphs should have a well-defined serialization

in order to convert data (obtained from a source system) into a syntax that can be stored (in the same system) or forwarded (to a target system), and then constructed again.

Data serialisation is very important for several reasons: it simplifies translation into other data formats; it enables automatic data processing; makes easier to compare databases, because the same data can be shared between systems; it is essential to foster the interoperability of heterogeneous databases; it results in a simpler backup method; other processing and visualization tools can read the data.

Our initial tasks consist of 1) surveying the state of the art, 2) understanding use cases and requirements for graph schemas, and 3) proposing a new property graph serialization that supports our requirements.

Graph serialization is nothing new. The lack of a specific Property Graph data model directly affects the development of other components, including query languages and serialization formats. Although there are some graph data syntaxes available (like GraphML or GEXF), there is no a standard one, and none of them is able to cover all the features presented by the property graph data model.

Although graph databases have gained popularity in the last years, their history goes many decades back. It is important to survey academic literature on graph serializations but also document-oriented database serializations. Academic results can significantly affect the scope of what can or should be done, and what cannot or should not be done.

Understanding the use cases and requirements of different people will help define the needs and the priority of features. Finally, on fundamentals of this, we propose YARS-PG that is platform independent and extensible serialization. Our serialization supports all the features allowed by the current database systems based on the property graph data model, and can be adapted to work with various visualization software, database-driven systems and other graph-oriented tools. In addition to grammar, we have also prepared ready-to-use parsers in Java, C# and Python.