From: **Alex Miłowski** alex@milowski.com
Subject: Position Statement
Date: 11 January 2019 at 19:50
To: group-data-ws-pc@w3.org

I am submitting this a my position statement for "W3C Workshop on Web Standardization for Graph Data". While currently an employee of Orange via Orange Silicon Valley, this position statement is my own.

Partitioning Graphs within Query Languages

Many of my current applications that utilize graph data are geo-spatially or time partitioned. While the geospatial regions of interest are fairly consistent, the time-series aspects are variable and depending upon what kind of analysis is being conducted for machine-learning or other analytics purposes. A typical query involves following a particular relationship in the graph from a subject of interest (e.g., a transaction) for a particular limit of depth for a particular time period. This kind of query and its resultants many need to be executed over a large data sets that are ever expanding.

Currently, the RDF model provides limited ways to model relations with properties, For example, a transfers between parties requires at least another subject between them to hold the properties of a transfer. This creates complexity in the SPARQL queries that limit the ability to both follow the relation and respect other constraints resulting in the use of UNIONS and degrading performance.

In contrast, a Property Graphs can represent this particular situation more directly but lacks the open extensibility of the RDF model. Further, the Cypher query language lacks property typing making following constraints like "time period" much more difficult.

It would like to see a better way to model relations and subject with a graph to have different dimensions of annotations. This would allow both direct modeling of vertex or edge properties and possibly provide an extensible way to do so. This would then require enhancements to the query languages to support constraints or traversal.

Meanwhile, query languages need to be aware of partitions. First, for scale-out applications, we need to be able to identify time or geospatial partitions. This will allow databases to route queries to the right partitions within a cluster.

Secondly, we need need better ways to describe constraints so that we can follow relationships within the graph without having to repeat them in specific detail to relate the constraints those relationships. That is, I want to say queries like "the connected subgraph of transfers between parties, from this origin, within this period of time, for a specific depth" in a concise way that results in performant queries.

Thank you for your time and consideration

--
--Alex Miłowski
"The excellence of grammar as a guide is proportional to the paucity of the
inflexions, i.e. to the degree of analysis effected by the language
considered."

Bertrand Russell in a footnote of Principles of Mathematics