Semantic Publishing Benchmark: Virtuoso Experience

Orri Erling

Openlink Software

November 14, 2014

Incentives of Semantic Publishing Benchmark

- Structure awareness for RDF
- Optimizing compile times, characteristic sets, plans reuse
- Bushy hash plans useful, but increase compile times
- Special graph-replace operation: Compare old and new content, delete/insert the difference

Inference

- Subproperty/subclass
- Equally convenient via materialization or backward chaining
- Small overhead of materialization
- Some schema inference possible but impact on metric not large

Promises of Structure Awareness

If patterns that share *Subject* are
- represented as tables for dense single valued *Predicates*
- treated as a unit in compilation

then:
- less joining at execution
- less plans to go through in compilation
- selective inlining of literals possible

Scaling SPB

- Literals and text index take the most space
- Text index dominates bulk load
- Quads under 20% of allocated space
- Queries between log and linear
- Short (log) and long (linear) queries counted in different metrics, some refinement of the mix still needed