Semantic Publishing Benchmark

London

November 19th, 2013

Motivation

- Use semantic technology to automate several steps in the publication pipeline
- Semantical annotation of content
- Media sectors using semantic technologies : news, finance, scientific publications

Use-case

- Scenario involves a media organization that maintains a catalogue of meta-data for its :
  - Journalistic assets (articles, photos, videos, papers, books, etc.)
- A piece of meta-data is called Creative Work
- Semantic Publishing Benchmark simulates :
  - Consumption of RDF metadata (Creative Works)
  - Management of RDF metadata (Creative Works)

Benchmark Design - Requirements

- Storing and processing RDF data
- Loading data in RDF serialization formats : Turtle, N-Quads
- Storing and isolating data in separate RDF graphs

Benchmark Design – Requirements 2

- Supporting following SPARQL standards : SPARQL 1.1 Query, SPARQL 1.1 Update, SPARQL 1.1 Protocol
- Support for RDFS, in order to return correct results
- Support for the RL profile of Web Ontology Language (OWL2 RL) in order to pass the conformance test suite

Benchmark Design – operational phases

- Initial loading of Ontologies and reference datasets
- Generation of Creative Works
- Loading of Creative Works
- Warm-up
- Benchmark
- Conformance tests (OWL2 RL)

Benchmark Configuration

- Number of editorial / aggregation agents
- Size of generated dataset (triples)
- Location of SPARQL endpoint i.e. URI
- Time length of Warm-up and Benchmark phases
- Each operational phase can be enabled or disabled

Benchmark Configuration 2

- Query-mix
  - Distribution of editorial operations
  - Distribution of aggregate operations
- Data Generator
  - Allocation of about / mentions tags
  - Popularity of an entity

Input Data - Ontologies

- Ontologies – provided by the BBC
  - Core ontologies : e.g. core concepts (things, places, events), persons, provenance, creative work, etc.
  - Domain ontologies : e.g. sports, news
  - Conformance ontologies : a part of the conformance test

Input Data – Reference Datasets

- Collection of entities describing various domains
  - Sports domain : football teams, formula1 teams
  - Politics : persons
  - Geonames : geo-locations

Data Generation – The Creative Work

- The meta-data about entities from reference data sets
- Has properties :
  - Title, short title, description, thumbnail
  - Creation date / modification date
  - Primary topic
  - Audience type
  - About / Mentions

The Workloads

- Simultaneous execution of editorial and aggregation agents

- Editorial agents – simulate editorial work performed by journalists :
  - Insert
  - Update
  - Delete

The Workloads 2

- Aggregation agents – simulate retrieval operations performed by end-users by executing :
  - Aggregation queries
  - Search queries
  - Geo-spatial , Full-text search queries
  - Drill-down queries (geo-locations, time-range)

Results Metrics

- Operations rate
  - Editorial operations per second
  - Aggregate operations per second
- Verbose mode
  - MIN, MAX, AVG execution time for each query
- All executed queries and results a saved to log files

Experimental Results

- Used different dataset sizes : 10M, 50M, 100M triples
- Benchmarked: OWLIM 5.4, Virtuoso7 OpenSrc
- Attempts to benchmark StarDog and BigData are in progress
- Benchmark configuration :
  - editorial agents : 2, aggregation agents : 14
  - warm-up : 60 s, benchmark : 300 s

Experimental Results Sample

Seconds run : 300

Editorial:

2 agents

1965 inserts (avg : 215 ms, min : 79 ms, max : 1462 ms)

258 updates(avg : 437 ms, min : 248 ms, max : 1370 ms)

242 deletes (avg : 234 ms, min : 95 ms, max : 1420 ms)

2465 operations (1965 CW Inserts (0 failed), 258 CW Updates (0 failed), 242 CW Deletions (0 failed))

8.2167 average operations per second

…

Experimental Results Sample

…

Aggregation:

14 agents

2351 Q1 queries (avg : 700 ms, min : 5 ms, max : 2778 ms, 0 failed)
2400 Q2 queries (avg : 7 ms, min : 3 ms, max : 1065 ms, 0 failed)
2358 Q3 queries (avg : 252 ms, min : 5 ms, max : 1618 ms, 0 failed)
2357 Q4 queries (avg : 101 ms, min : 2 ms, max : 1436 ms, 0 failed)
2292 Q5 queries (avg : 57 ms, min : 3 ms, max : 1345 ms, 0 failed)
2381 Q6 queries (avg : 38 ms, min : 19 ms, max : 1260 ms, 0 failed)
2341 Q7 queries (avg : 601 ms, min : 5 ms, max : 2626 ms, 0 failed)

16480 total retrieval queries (0 failed)

54.9333 average queries per second

Experimental Results Summary

- Results for OWLIM and Virtuoso (reduced query-mix)

Dataset Size	OWLIM 5.4		Virtuoso 7 OpenSource	
	Ed. ops	Aggr. ops	Ed. ops	Aggr. ops
10 M	9.1	68.8	142.7	? 2.9
50 M	8.1	52.9	140.7	17.8
100 M	5.8	39.2	? 3.55	? 0.5

- Disclaimers: initial results before calibration
  - Virtuoso's geo-spatial indices not used when measuring the results above

Future Work

- Further fine-tuning of aggregate query-mix is necessary
- Validation of results
- Data generation – finding a balance between the amount of generated creative works and the reference data size

Questions

?