Business Intelligence Over Linked Open Spatio-Temporal Data

Nurefşan Gür
Aalborg University, Denmark
Supervisors: Torben Bach Pedersen & Esteban Zimányi & Katja Hose

What is My Research About

- Semantic Spatio-Temporal (ST) Warehouses
- RDF Data Warehousing (DW) with Business Intelligence (BI) technologies
- Geospatial Semantic Web (SW)
- Spatio-Temporal Linked Open Data (LOD).

Tools and Approaches:
- SW Tools: RDF Data Model, SPARQL 1.1, GeoSPARQL*, Triple Stores
- BI Tools: Spatial Data Warehouses, OLAP Cubes, MDX
- Query Processing approaches: Parallel, Federated over Distributed RDF Stores

*Note that possible future direction of tools and use case data written in italic grey fonts

Overall Objectives & Expected Outcomes

1. To represent data with multi-dimensional, enhanced ontologies and integrate it into a knowledge base linked with relevant data.
2. To support advanced link discovery for spatial objects and to develop advanced query techniques for multi-dimensional RDF data.
3. To extend query processing capabilities over distributed RDF data sources and test the applicability with OLAP queries

Proof-of-concept RDF Data Model and Ontology Design Pattern (ODP) for ST data
Spatial Link/Relation Discovery Tool based on Region Connection Calculus (RCC8)+ ST Query Language
OLAP Queries and results over distributed RDF data sources

Work Flow

1. Re-process in the cycle steps A, B and possibly D to link external sources i.e. DBpedia, Freebase etc. (see Figure 1.)
2. Advanced linking process (D): test on spatial objects with RCC8 predicates (intersects, touches, overlaps) & Extendend SPARQL for ST Data
3. Explore and derive Bi from chosen domains & present a demo interface for decision makers

Process Cycle

Queries

SELECT ?crop COUNT(*) AS ?cnt
FROM <http://extbi.lab.aau.dk/resource/agriculture>
WHERE {
?crop a dcterms:hasEdition .
FILTER (str(?coverage) = "Europe")
GROUP BY ?crop
}

Query 1: AQT† - Counts fields based on the crop they produce

SELECT ?name ?address
FROM <http://extbi.cs.aau.dk/resource/agriculture>
WHERE {
?company owl:sameAs [a ?type] .
?company rdfs:comment ?address
}

Query 2: SQT* - Finds the company address that owns organic fields

Use Case

Agricultural Domain
- Field
- Organic Field
- Field Block
- Livestock Farming
- Seasonal Farming
- Annual Expenditures

Climate Sensor Data
- Precipitation sensors
- Wind Turbine sensors

Business Domain
- Company Registry Data (CVR)
- Production units & addresses
- Fully responsible participant data

Geographical data
- Municipalities
- Fields
- Water bodies

Evaluation

<table>
<thead>
<tr>
<th>Step</th>
<th>Virtual</th>
<th>Materialized</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Cleaning</td>
<td>74.92</td>
<td>603.35</td>
<td>603.35</td>
</tr>
<tr>
<td>Load Ontology</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Load Mappings</td>
<td>8.76</td>
<td>12.35</td>
<td>12.35</td>
</tr>
<tr>
<td>Dump RDF</td>
<td>0.00</td>
<td>4684.82</td>
<td></td>
</tr>
<tr>
<td>Load RDF</td>
<td>0.00</td>
<td>840.04</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>84.68</td>
<td>616.70</td>
<td>614.56</td>
</tr>
</tbody>
</table>

Table 1: Load times in seconds

<table>
<thead>
<tr>
<th>Query</th>
<th>Virtual</th>
<th>Materialized</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQT 1</td>
<td>5.92</td>
<td>3.39</td>
<td>1.04</td>
</tr>
<tr>
<td>AQT 2</td>
<td>13.32</td>
<td>7.00</td>
<td>0.23</td>
</tr>
<tr>
<td>AQT 3</td>
<td>10.81</td>
<td>7.70</td>
<td>0.05</td>
</tr>
<tr>
<td>AQT 4</td>
<td>-</td>
<td>-</td>
<td>0.14</td>
</tr>
<tr>
<td>AQT 5</td>
<td>-</td>
<td>20.37</td>
<td>0.86</td>
</tr>
<tr>
<td>SQT 1</td>
<td>-</td>
<td>-</td>
<td>2.35</td>
</tr>
<tr>
<td>SQT 2</td>
<td>0.09</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>SQT 3</td>
<td>2188.85</td>
<td>1.81</td>
<td>0.40</td>
</tr>
<tr>
<td>SQT 4</td>
<td>6.57</td>
<td>2.35</td>
<td>1.63</td>
</tr>
<tr>
<td>SQT 5</td>
<td>-</td>
<td>23.79</td>
<td>3.29</td>
</tr>
<tr>
<td>Average</td>
<td>370.93</td>
<td>8.31</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Table 2: Runtimes in seconds

*AQT : Aggregated Query Template / SQT: Simple Query Template

References

1 M3 - https://it4bi-db.aub.ac.uk/Modeling_and_Semantics
2 Extended BI: EXTBi Project page: http://extbi.lab.aau.dk/