Implemented RDF Stores with Geospatial Support

Presenter: Kostis Kyzirakos
Outline

- Relational DBMS with a geospatial extension

- RDF stores with a geospatial component:
  - Research prototypes
  - Commercial systems
Geometries are not explicitly handled by query language (SQL)

Define datatypes that extend the SQL type system

- Model geometries using Abstract Data Type (ADT)
- Hide the structure of the data type to the user
  - The interface to an ADT is a list of operations
    - For spatial ADTs: Operations defined according to OGC Simple Features for SQL
  - Vendor-specific implementation irrelevant - extend SQL with geometric functionality independently of a specific representation/implementation
How does an RDBMS handle geometries? (2/2)

Special indices needed for geometry data types
Specialised query processing methods
Implemented Systems

Will examine following aspects:

- Data model
- Query language
- Functionality exposed
- Coordinate Reference System support
- Indexing Mechanisms
Research Prototypes

- Strabon
- Parliament
- Brodt et al.
- Perry
Strabon

- Storage and query evaluation module for stSPARQL
- Geometries represented using typed literals
  - WKT & GML serializations supported
- Spatial predicates represented as SPARQL functions
  - OGC-SFA, Egenhofer, RCC-8 families exposed
  - Spatial aggregate functions
- Support for multiple coordinate reference systems

- GeoSPARQL support
  - Core
  - Geometry Extension
  - Geometry Topology Extension
Strabon - Implementation

Parliament

- Storage Engine
- Developed by Raytheon BBN Technologies
- Implementation of GeoSPARQL
  - Geometries represented using typed literals
    WKT & GML serializations supported
  - Three families of topological functions exposed
    OGC-SFA
    Egenhofer
    RCC-8
  - Multiple CRS support

[Battle and Kolas, 2011]
Parliament - Implementation

- Rule engine included
- Paired with query processor
- R-tree used

Brodt et al.

- Built on top of RDF-3X
- Implemented at University of Stuttgart
- No formal definitions of data model and query language given
- Geometries expressed according to OGC-SFA
  - Typed Literals
  - WKT serialization supported
  - Expressed in WGS84
- Spatial predicates represented as SPARQL filter functions
  - OGC-SFA functionality exposed

[Brodt et al., 2010]
Focus on spatial query processing and spatial indexing techniques for spatial selections

e.g. "Retrieve features located inside a given polygon"

Naive spatial selection operator

Placed in front of the execution plan which the planner returns

Spatial index (R-Tree) implemented

Only utilized in spatial selections

Available upon request
- Built on top of Oracle 10g
- Implemented at Wright State University
- Implementation of SPARQL-ST
  - Upper-level ontology imposed
- Geometries expressed according to GeoRSS GML
- Spatial and temporal variables introduced
- Spatial and temporal filters used to filter results with spatiotemporal constraints
  - RCC-8 calculus
  - Allen’s interval calculus
Perry - Implementation

- Spatiotemporal operators implemented using Oracle's extensibility framework
  - Three spatial operators defined
- Strictly RDF concepts implemented using Oracle’s RDF storage and inferencing capabilities
- R-Tree used for indexing spatial objects

Available upon request
Commercial RDF Stores

- AllegroGraph
- OWLIM
- Virtuoso
- uSeekM
AllegroGraph

- Well-known RDF store, developed by Franz Inc.
- Two-dimensional point geometries
  - Cartesian / spherical coordinate systems supported
- GEO operator introduced for querying
  - Syntax similar to SPARQL’s GRAPH operator
  - Available operations:
    - Radius / Haversine (Buffer)
    - Bounding Box
    - Distance

- Linear Representation of data
  - X and Y ordinates of a point are combined into a single datum
- Distribution sweeping technique used for indexing
  - Strip-based index
OWLIM

- Semantic Repository, developed by Ontotext
- Two-dimensional point geometries supported
  Expressed using W3C Geo Vocabulary
    Point Geometries
    WGS84
- Spatial predicates represented as property functions
  Available operations:
    Point-in-polygon
    Buffer
    Distance

- Implemented as a Storage and Inference Layer for Sesame
- Custom spatial index used
- Closed Source
  Free version available for evaluation purposes
  http://www.ontotext.com/owlim
Virtuoso

- Multi-model data server, developed by OpenLink
- Two-dimensional point geometries
  - Typed literals
  - WKT serialization supported
  - Multiple CRS support
- Spatial predicates represented as functions
  - Subset of SQL/MM supported

- R-Tree used for indexing
- Spatial capabilities firstly included in Virtuoso 6.1
- Closed Source

Open Source Edition available from
http://virtuoso.openlinksw.com/

Does not include the spatial capabilities extension
Add-on library for Sesame-enabled semantic repositories, developed by OpenSahara

Geometries expressed according to OGC-SFA
  - WKT serialization
  - Only WGS84 supported

Spatial predicates represented as functions
  - OGC-SFA functionality exposed
  - Additional functions
    - e.g. shortestline(geometry,geometry)

Implemented as a Storage and Inference Layer (SAIL) for Sesame
  - May be used with RDF stores that have a Sesame Repository/SAIL layer

R-tree-over-GiST index used (provided by PostGIS)

Open Source, Apache v2 License
  - Available from https://dev.opensahara.com/projects/useekm
<table>
<thead>
<tr>
<th>System</th>
<th>Language</th>
<th>Index</th>
<th>Geometries</th>
<th>CRS support</th>
<th>Comments on Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strabon</td>
<td>stSPARQL/GeoSPARQL*</td>
<td>R-tree-over-GiST</td>
<td>WKT / GML support</td>
<td>Yes</td>
<td>• OGC-SFA • Egenhofer • RCC-8</td>
</tr>
<tr>
<td>Parliament</td>
<td>GeoSPARQL</td>
<td>R-Tree</td>
<td>WKT / GML support</td>
<td>Yes</td>
<td>• OGC-SFA • Egenhofer • RCC-8</td>
</tr>
<tr>
<td>Brodt et al. (RDF-3X)</td>
<td>SPARQL</td>
<td>R-Tree</td>
<td>WKT support</td>
<td>No</td>
<td>OGC-SFA</td>
</tr>
<tr>
<td>Perry</td>
<td>SPARQL-ST</td>
<td>R-Tree</td>
<td>GeoRSS GML</td>
<td>Yes</td>
<td>RCC-8</td>
</tr>
<tr>
<td>AllegroGraph</td>
<td>Extended SPARQL</td>
<td>Distribution</td>
<td>2D point geometries</td>
<td>Partial</td>
<td>• Buffer • Bounding Box • Distance</td>
</tr>
<tr>
<td>OWLIM</td>
<td>Extended SPARQL</td>
<td>Custom</td>
<td>2D point geometries</td>
<td>No</td>
<td>• Point-in-polygon • Buffer • Distance</td>
</tr>
<tr>
<td>Virtuoso</td>
<td>SPARQL</td>
<td>R-Tree</td>
<td>2D point geometries (in WKT)</td>
<td>Yes</td>
<td>SQL/MM (subset)</td>
</tr>
<tr>
<td>uSeekM</td>
<td>SPARQL</td>
<td>R-tree-over-GiST</td>
<td>WKT support</td>
<td>No</td>
<td>OGC-SFA</td>
</tr>
</tbody>
</table>
Conclusions

- **Semantic Geospatial Systems:**
  - Research Prototypes
  - Commercial Systems

- **Next topic:** Geospatial information with description logics, OWL and rules
Bibliography

[Kyzirakos et al, 2010]

[Kyzirakos et al, 2012]

[Battle and Kolas, 2011]
Bibliography

[Brodt et al, 2010]

[Perry, 2007]