Implemented RDF Stores with Geospatial Support

Presenter: Kostis Kyzirakos

Dept. of Informatics and Telecommunications
National and Kapodistrian University of Athens

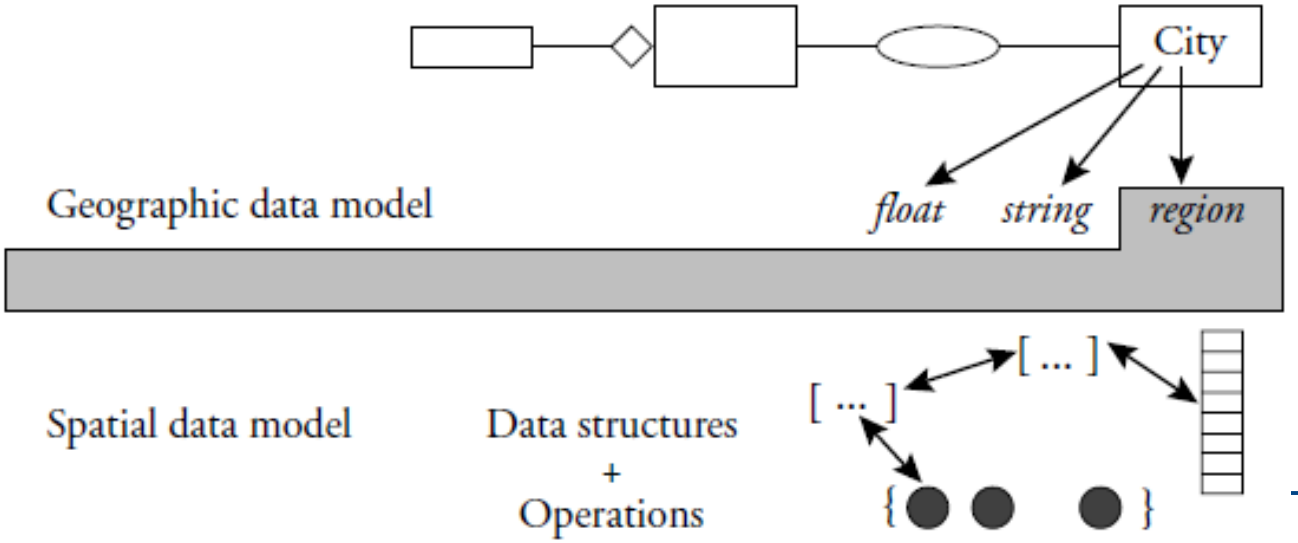
Outline

- Relational DBMS with a geospatial extension

- RDF stores with a geospatial component:
  - Research prototypes
  - Commercial systems

How does an RDBMS handle geometries? (1/2)

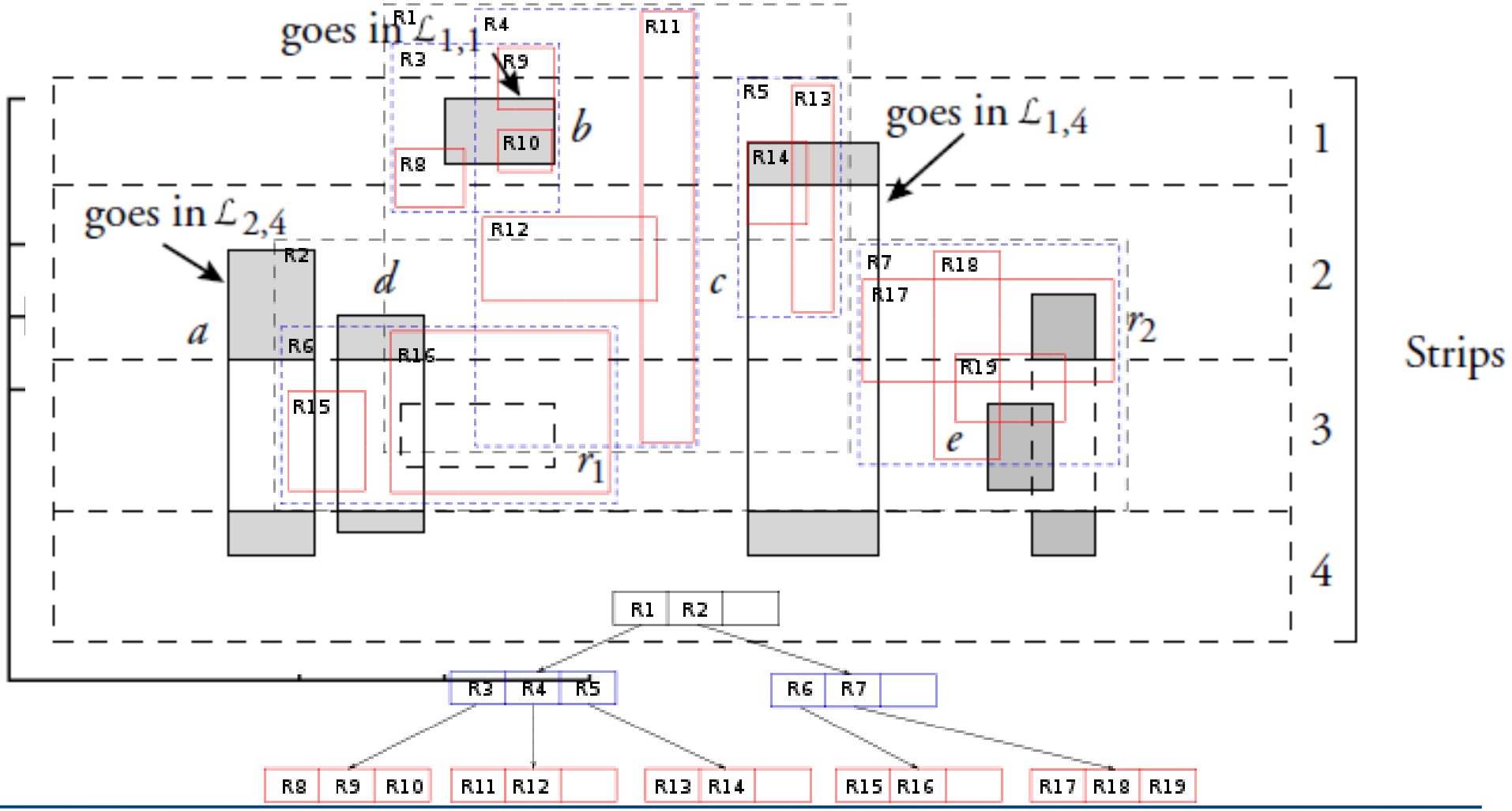
- Geometries are not explicitly handled by query language (SQL)
- Define datatypes that extend the SQL type system
  - Model geometries using Abstract Data Type (ADT)
  - Hide the structure of the data type to the user
    - The interface to an ADT is a list of operations
      - For spatial ADTs: Operations 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

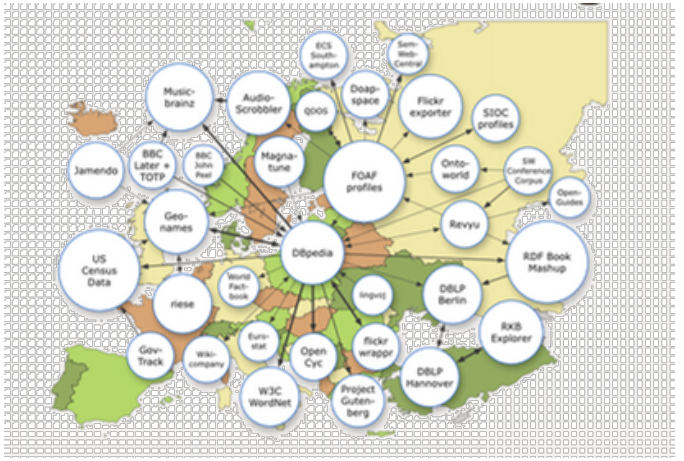
- Strabon
- Parliament
- Brodt et al.
- Perry

Strabon

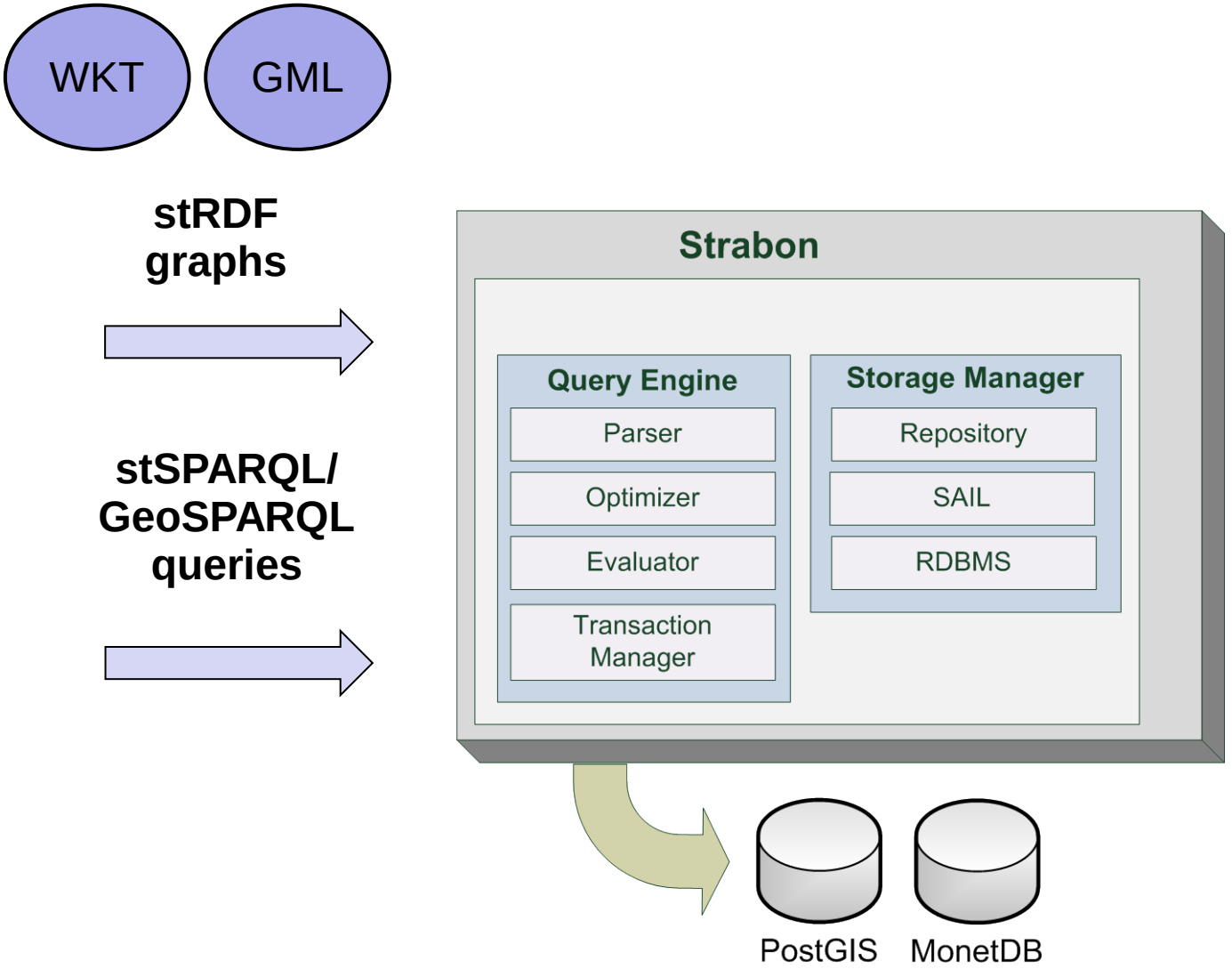
- Storage and query evaluation module for stSPARQL

[Kyzirakos et al.,'10,'12]

- 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



Open Source, available from http://www.strabon.di.uoa.gr/

Parliament

- Storage Engine [Battle and Kolas, 2011]
- Developed by Raytheon BBN Technologies
- Implementation of GeoSPARQL
  - Geometries represented using typed literals

    WKT & GML serializations supported
  - Three families of topological functions exposed

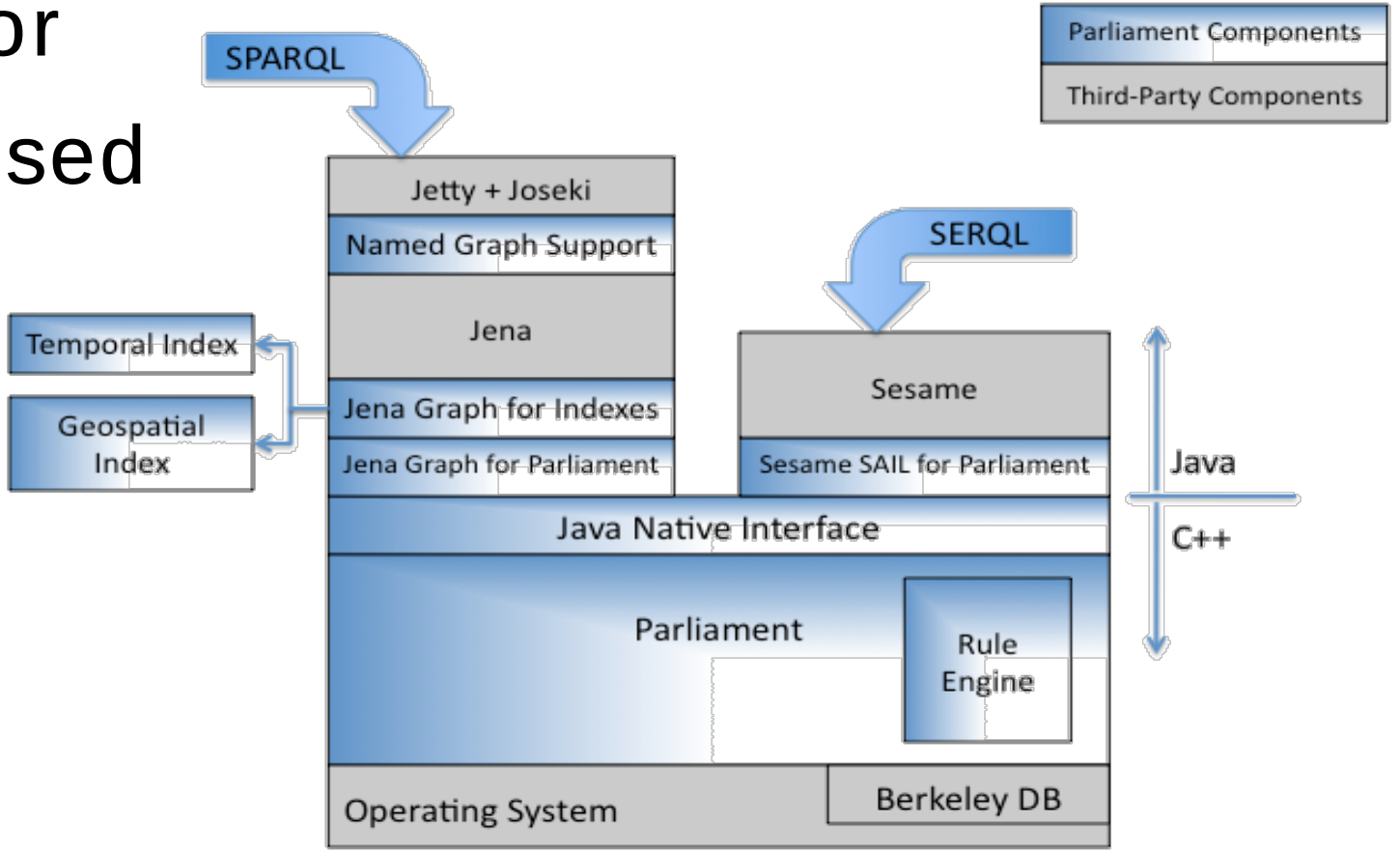
    OGC-SFA

    Egenhofer

    RCC-8
  - Multiple CRS support

Parliament - Implementation

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



Open Source, available from
http://www.parliament.semwebcentral.org

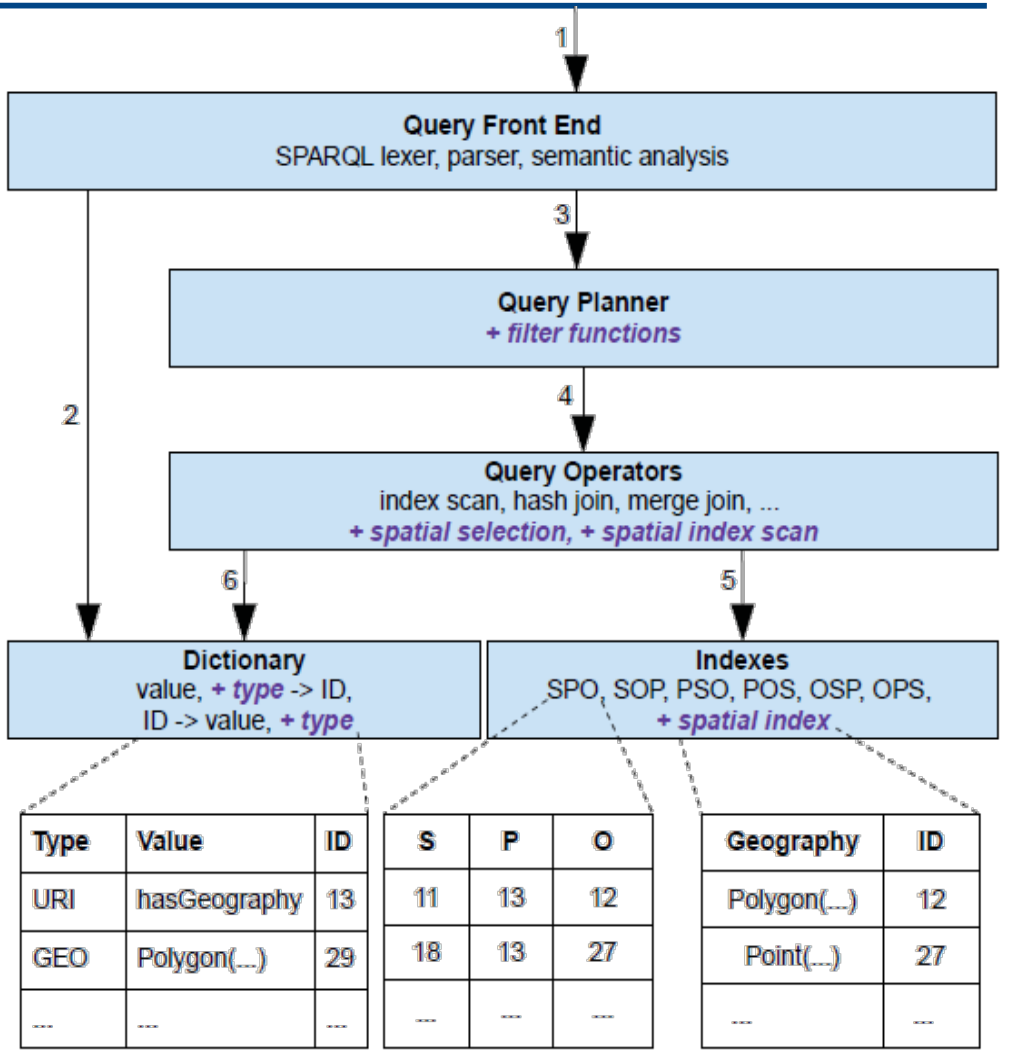
Brodt et al.

[Brodt et al., 2010]

- 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. - Implementation

- 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

Perry

- Built on top of Oracle 10g [Perry, 2008]
- Implemented at Wright State University
- Implementation of SPARQL-ST

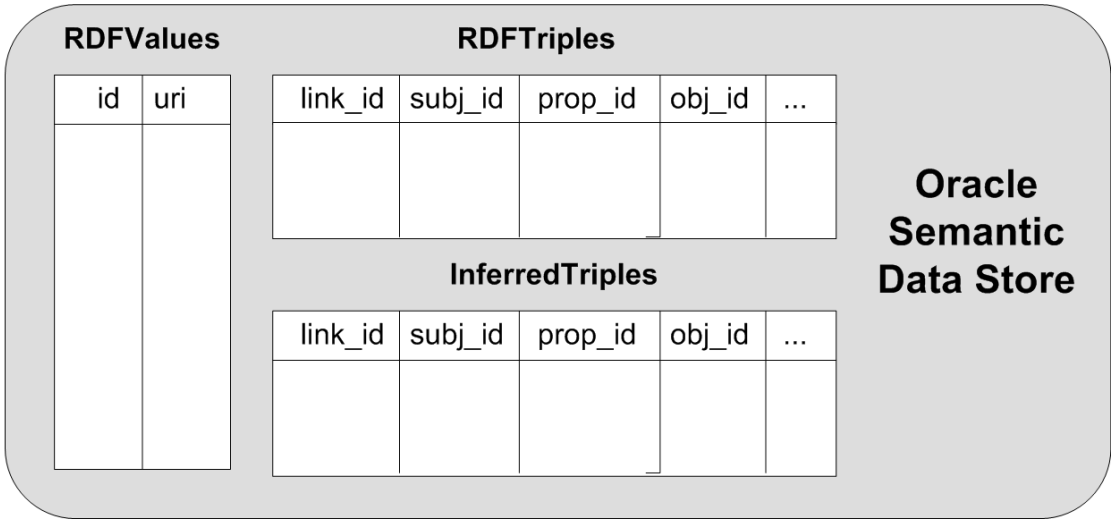
  Upper-level ontology imposed
- Geometries expressed according to GeoRSS GML
- Spatial and temporal variables introduced
- Spatial and temporal filters used to filter results with spatiotemporal constraints

  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



SpatialData

value_id	shape

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
- Closed source, available from http://www.franz.com/agraph/allegrograph/

OWLIM

ontotext

- 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

uSeekM

- 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

System	Language	Index	Geometries	CRS support	Comments on Functionality
Strabon	stSPARQL/ GeoSPARQL*	R-tree-over-GiST	WKT / GML support	Yes	• OGC-SFA • Egenhofer • RCC-8
Parliament	GeoSPARQL	R-Tree	WKT / GML support	Yes	•OGC-SFA •Egenhofer •RCC-8
Brodt et al. (RDF-3X)	SPARQL	R-Tree	WKT support	No	OGC-SFA
Perry	SPARQL-ST	R-Tree	GeoRSS GML	Yes	RCC-8
AllegroGraph	Extended SPARQL	Distribution sweeping technique	2D point geometries	Partial	•Buffer •Bounding Box •Distance
OWLIM	Extended SPARQL	Custom	2D point geometries (W3C Basic Geo Vocabulary)	No	•Point-in-polygon •Buffer •Distance
Virtuoso	SPARQL	R-Tree	2D point geometries (in WKT)	Yes	SQL/MM (subset)
uSeekM	SPARQL	R-tree-over GiST	WKT support	No	OGC-SFA

Conclusions

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

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

Bibliography

[Kyzirakos et al, 2010]
K. Kyzirakos , M. Karpathiotakis, M. Koubarakis: Developing Registries for the Semantic Sensor Web using stRDF and stSPARQL (short paper). In: Proceedings of the 3rd International Workshop on Semantic Sensor Networks (SSN10) (2010)

[Kyzirakos et al, 2012]
K. Kyzirakos , M. Karpathiotakis, M. Koubarakis: *Strabon: A Semantic Geospatial DBMS*. In: Proceedings of the 11th International Semantic Web Conference (2012)

[Battle and Kolas, 2011]
Battle, R., Kolas, D.: *Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL* (2011)

Bibliography

[Brodt et al, 2010]

A. Brodt, D. Nicklas, and B. Mitschang. *Deep integration of spatial query processing into native rdf triple stores*. In ACM SIGSPATIAL, 2010.

[Perry, 2007]

Matthew Perry. *A Framework to Support Spatial, Temporal and Thematic Analytics over Semantic Web Data*. PhD thesis, Wright State University, 2008