RDF Data Model Tutorial

Susie Stephens
Principal Product Manager

Bill Gettys
interMedia Product Manager
Outline

- RDF Data Model Overview
- Family Tree Use Case
- RDF & Image Use Case
- Protein Use Case
Oracle RDF Data Model
Oracle RDF Data Model

- Support for RDF and RDFS
- Object-relational implementation
- Subjects and objects are re-used
- Links represent complete RDF triples

RDF Triples:
- \{S_1, P_1, O_1\}
- \{S_1, P_2, O_2\}
- \{S_2, P_2, O_2\}
SPARQL-like Query Capability

- A table function allows a graph query to be embedded in a SQL query
- Searches for an arbitrary pattern against the RDF data
- Includes inferencing based on RDF, RDFS, and user-defined rules
Enterprise Functionality

- Real Application Clusters (RAC), Security
- Multi-threaded, parallel processing, indexed, etc.
- Integrated with other data types in the database
- Performance testing with UniProt

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 M Triples</td>
<td>0.86</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.18</td>
<td>0.46</td>
</tr>
<tr>
<td>20 M Triples</td>
<td>0.95</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.19</td>
<td>0.47</td>
</tr>
<tr>
<td>40 M Triples</td>
<td>0.96</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.18</td>
<td>0.47</td>
</tr>
<tr>
<td>80 M Triples</td>
<td>1.03</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.20</td>
<td>0.49</td>
</tr>
<tr>
<td>Maximum σ</td>
<td>0.054</td>
<td>0.002</td>
<td>0.002</td>
<td>0.011</td>
<td>0.065</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Units in seconds

Source: Chong et al. VLDB 2005
RDF and Image Use Case
Use Case Components

- JSP application
- DICOM images
- RDF schema
- Query data though the Web interface
Image Search

- Images can be searched using metadata
  - DICOM, TIFF, GIF, JPEG have metadata as part of the image
  - Metadata extracted from the images and stored in table columns
  - Metadata can be stored as XML enabling XML based query
Image Search

“Find me all DICOM images that contain the term ‘Jaw’”

1. Keyword search
Image Search

“Find me all DICOM images that contain the term ‘Jaw’”

1. Keyword search
2. Map relationships to terms using RDF triples
SNOMED Relationships

- Maxilla
  - T-11170

- Mandible
  - T-11180

- Maxilla and Mandible
  - T-D1217
... more

T-D1217

Includes
T-11180

Includes
T-11170
... and

Jaw

Equals

Maxilla and Mandible
Family Tree Use Case
Use Case Components

- Load family tree data
- Create a rule base
- Create a rule
- Index the family tree data
- Perform a query with inferencing
- Perform a query without inferencing
Embedding RDF Query in SQL

SELECT ...
FROM ..., TABLE ( RDF Query
(expressed via SDO_RDF_MATCH invocation) ) t, ...
WHERE ...
SDO_RDF_MATCH

• Input Parameters
  
  SDO_RDF_MATCH ( 
  Query, \rightarrow \text{graph-pattern (with variables)}  
  Models, \rightarrow \text{set of RDF models}  
  Rulebases, \rightarrow \text{set of rulebases (e.g., RDFS)}  
  Aliases, \rightarrow \text{aliases for namespaces}  
  Filter \rightarrow \text{additional selection criteria}  
  )  

• Return type in definition is AnyDataSet  
• Actual return type is determined at compile time based on the arguments for each specific invocation
Rulebase: Overview

- Each rulebase consists of a set of rules
- Each rule consists of
  - antecedent: graph-pattern
  - filter condition (optional)
  - consequent: graph-pattern
- One or more rulebases may be used with relevant RDF models (graphs) to infer new data
Rules Index: Overview

- A rules index is created on an RDF dataset (consisting of a set of RDF models and a set of RDF rulebases)
- A rules index contains RDF triples inferred from the model-rulebase combination
Protein Use Case
Use Case Components

- Biocyc and Reactome have been loaded into the RDF Data Model
- Perform more complex graph queries over the data
Map Pathway to BioPAX Format

Summary

- Oracle has a scalable, secure, highly-available RDF Data Model
- Oracle has extended SQL to perform graph pattern matching (with inferencing)
- SQL queries can be used to link RDF data with multiple data types