RDF in the Database: Enabling Semantically Rich Business Applications

Xavier Lopez
Director, Product Management
Server Technologies
Overview

- Semantic Applications Opportunity
- Semantic Technology Primer - RDF
- Germane Oracle10g Database Technologies
  - Network Data Model
  - RDF
  - OWL
Networks to Model Systems
Life Sciences, Ontologies, Social Networks
Network Concepts:

A network is a graph representation for modeling objects of interest and their relationships. It contains the following elements:

- **Nodes**: objects of interests (e.g. intersection, subject)
- **Links**: relationship between nodes (e.g. street, property)
- **Paths**: ordered list of connected links (e.g. scenic route, interaction)

- RDF introduces a universal way to manage graph representation
RDF Data Model Opportunities

- **Unique Business Opportunities**
  - Life Sciences: pathway analysis, protein interaction
  - Web: service discovery, FOAFs, blogs
  - eBusiness: grid resources, app integration, BI
  - Security: social networks, provence, varying trust

- **Applying DBMS Technology to the Challenge**
  - Scalability: models comprising millions of graphs
  - Security: Web-based, trust, reification
  - Transaction, versioning, performance
  - Exploit expressive power of SQL
  - Interoperability: Integrating multiple networks
Semantic Technologies

**Goal**: Associating more meaning (context) to enterprise data to enable its use across applications
- Common framework to express information to be exchanged between applications without loss of meaning

**How?** Develop technologies to allow sharing and reuse of enterprise and web data.
W3C Stack of languages

- **XML**: Surface syntax, no semantics
- **XML Schema**: Describes structure of XML documents
- **RDF**: Graph data model for “relations” between “things”
- **RDF-Schema**: Defines a domain vocabulary for RDF
- **OWL**: A more expressive Vocabulary Definition Language
Resource Description Framework (RDF)

• Originally conceived as W3C’s metadata model
  • Document metadata for digital libraries, content rating, site maps, etc.

• Simple data model
  • Leverages syntactic extensibility and modularity of XML namespaces
  • Provides global extensibility through a common data model
  • Directed labeled graph: “subject/property/object”
  • Nodes are called “resources” and links “properties”
RDF Statements (1)

- RDF specifies simple descriptions of resources
- Subject -&gt; Property -&gt; Object triples
- Objects are web-resources (URIs)
- Subject is again an Object:
  - triples can be linked
RDF Statements (2)

- Every identifier is a URI
  - = world-wide unique naming!

- Any statement can be a resource
  - represented by a node
  - graphs can be nested
  - Reification example:
    - Metadata: author, date, source.
RDF Schema (RDFS)

- Allows creation of class and property hierarchies
  - Ex: `<\`NCI:Rheumatoid_Arthritis``
    \`rdfs:subClassOf``
    \`NCI:Autoimmune_Disease`'>
  - Defines a domain vocabulary for RDF
  - Organises this vocabulary into a typed hierarchy
- RDF Schemas have URIs and can be described using RDF
RDFS Data Model
RDF Inferencing

- Employing symmetry and transitivity characteristics of properties to infer new relationships
- RDF Statements + RDFS rules
- Syntax for specifying user-defined rules
  - Enabled by RDFS
Beyond RDF: OWL
Web Ontology Language

- OWL builds on RDF and RDF Schema for describing sophisticated relationships for inclusion in (domain-specific) Ontologies
- Incorporates Ontologies being developed by various industry domains
  - Life Sciences
  - National Security
  - Health Care
  - Manufacturing
Oracle10g Network Data Model
Oracle Spatial NDM

☑ Provide an open and generic network data model and analysis platform for graph-based applications (store, index, query)
☑ Combine specialized application information with a general network data
☑ Applies efficient network algorithms and constraints to support graph analysis
☑ Enable 3rd party tools and apps
NDM Analytical Capabilities

- Canonical network analysis functions
  - Shortest path between two nodes
  - All paths between two nodes
  - Nodes reachable from a given node
  - All nodes within cost $c$ from a given node
  - Nodes capable of reaching a given node
  - Nearest $n$ nodes from a given node
  - Minimum Cost Spanning Tree
  - Traveling Salesman Problem

- Framework to apply network constraints
  - (path length, cost, links to avoid, etc.)
Perform Shortest Path Analysis in 10g
Shortest Path Analysis: KEGG Data
NDM Architecture

Thick or Thin Client
Browsing, navigation, presentation, editing, and analysis

Java API:
Network features loaded as in-memory Java objects in client tier or middle tier

Network Schema:
Persistent node, link, path and path-link tables along with metadata

Oracle Spatial 10g
RDF Support in Oracle10g R2
RDF Design Objectives:

- Extend NDM to support RDF object types
- Enable RDF with graph analysis
- Enable combined SQL query of enterprise database and RDF graphs
- Support large, complex graphs (10s of millions statements)
- Easily extensible by 3rd party tools/apps
RDF in Oracle Spatial

- RDF data stored in a directed, logical network
- **Subjects** and **objects** mapped to nodes, and **properties** to links that have subject start nodes and object end nodes
- 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\}
Two new RDF object types
- SDO_RDF_TRIPLE
- SDO_RDF_TRIPLE_S

Several constructors and member functions

RDF Data Types (URIs, Blank Nodes, Plain, Typed and Long Litterals, Collections Types)

Multiple representations of values

Reification
RDF Querying: Overview

- Extend SQL using Extensibility Framework
  - RDF_MATCH table function
  - for graph pattern specification in query
- SPARQL –like syntax (W3C) support
- SQL-level access to RDF data
- Java API
RDF Querying: Example

Find pairs of persons residing at the same address such that first person rents a truck and the second person buys fertilizer

```
SELECT t3.x name1, t3.y name2
FROM AddrTable t1, AddrTable t2,
TABLE(RDF_MATCH(RDFModels('transactions'), ...
  `(?x :rents ?a) (?a rdf:type :truck)
  (?y :buys ?b) (?b rdf:type :fertilizer)`
  ...)) t3
WHERE t1.name=t3.x and t2.name=t3.y and
  t1.addr=t2.addr;
```

Shows embedding of a RDF query in a SQL query
RDF Querying: Example

Find pairs of persons residing at the same address such that first person rents a truck and the second person buys fertilizer

```
SELECT t3.x name1, t3.y name2
FROM AddrTable t1, AddrTable t2,
TABLE(RDF_MATCH(
    '(?x :rents ?a) (?a rdf:type :truck)
    (?y :buys ?b) (?b rdf:type :fertilizer)',
    RDFModels('activities','vehicles','chemicals'),
    RDFRulebases('rdfs')
    ...)) t3
WHERE t1.name=t3.x and t2.name=t3.y and
    t1.addr=t2.addr;
```

Shows embedding of a graph query in a SQL query
Use Cases:

- Astra Zeneca – Biological Pathways Analysis
- Beyond Genomics – Pathways Analysis
- GlaxoSmithKlein – Pathways Analysis
- Kyoto Univ. (KEGG) – Pathways Analysis
- Nature Publishing – Citation Analysis
- Overstock.com – Customer Care
- Siderean – Enterprise Search Partner
- Tom Sawyer – Graph Visualization Tool
- Cytoscape – Graph Visualization Tool
Planned OWL Support

Oracle’s Approach:

- Allow OWL Ontologies to be stored in a set of Oracle tables
- Provide a collection of SQL Operators and SQL Table Functions to query OWL Ontologies
## Restaurant Example

### served_food

<table>
<thead>
<tr>
<th>R_id</th>
<th>Cuisine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>American</td>
</tr>
<tr>
<td>2</td>
<td>Mexican</td>
</tr>
<tr>
<td>2</td>
<td>American</td>
</tr>
<tr>
<td>14</td>
<td>Brazilian</td>
</tr>
</tbody>
</table>

### Cuisine_ontology

- **cuisine**
  - Latin American
  - Mexican
  - Brazilian

...
Querying Cuisine Ontology

```
SELECT * FROM served_food
WHERE ONT_RELATED ( cuisine,
   'IS_A',
   'Latin American',
   'Cuisine_ontology')=1;
```

Result:

<table>
<thead>
<tr>
<th>R_id</th>
<th>Cuisine</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Mexican</td>
</tr>
<tr>
<td>14</td>
<td>Brazilian</td>
</tr>
</tbody>
</table>
Additional Ontology Queries

- Semantic filtering
- Query with path length constraints
- Query with path content constraints
- Semantic join
Status

Design & Implementation

- RDF Querying → 10gR2 database
- OWL Querying → prototyped, planned for 11x database