Triple Stores: Jena

Luke Georgalas
CMSC 691S

Agenda

- Quick introduction to triple stores
- Jena environment
- Jena features
- Under the hood
- Sample Jena code and demo: importing data and querying

Quick Introduction To Triple Stores

- Knowledge is represented as subject, predicate, object
- Triple stores store this knowledge in a database
- Most triple stores allow you to store the model in memory, or persist the data to disk in a number of different ways
- Most support both listing all knowledge and using SPARQL queries
- Some include reasoning engines, or hooks to use use external reasoning engines
- Quad stores also exist, which allow you to store the source of a triple

Jena Environment

- Jena is a Java library
- OS-independent
- Supports external databases through JDBC
  - MySQL
  - PostGreSQL
  - SQL Server (Microsoft)
  - HSQLDB
  - Oracle
  - DB2 (IBM)
  - Derby (Apache)
- Supports external reasoning engines through DIG
  - Any DIG-compliant reasoning engine
  - Pellet has a special binding

Jena Features - Input/Output

- Supported formats
  - RDF/XML
  - N3
  - N-Triple
  - Turtle
- Input
  - STDIN
  - Local files
  - Remote files
  - Programmatic manipulation
- Output
  - Local files
- Command line interface

Jena Features - Models

- Read/Write models
  - In Memory
  - Database
  - File
- Other sources of data that may present triples
  - Filesystem
  - Other implementation of the model interface

Jena Features - Querying

- Programmatic access to models
  - List statements
  - Combine models and schemas
- SPARQL

Jena Features - Inference Engine

- Built in inference engine
  - RDF, RDFS, OWL-Lite (partially supported), DAML+OIL
- Hooks to use DIG-compliant external inference engines
- Pellet has a special binding for improved performance

Under The Hood - Enhanced Graph Layer

- Jena does not attempt to map to a Java object model
- Object polymorphism is supported
- Model API and Ontology Model API

Under The Hood - Performance Considerations

- Fast Path Queries
  - Allow underlying data store to interpret queries
- Denormalization
  - Space/speed tradeoff in database storage
- Models can be copied to different stores using API

Demo

- Loading and printing content
- Inference engine - OWL
- Persistence
- SPARQL query

References

- Jena web site
  - http://jena.sourceforge.net/documentation.html
- Wikipedia
  - http://en.wikipedia.org/wiki/Jena_(framework)
- Jeremy J. Carroll, Ian Dickinson, Chris Dollin, Dave Reynolds, Andy Seaborne, Kevin Wilkinson, Jena: implementing the semantic web recommendations, Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, May 19-21, 2004, New York, NY, USA
  - http://cs.umbc.edu/691s/papers/jena.pdf