CIS 602: Provenance & Scientific Data Management

Provenance & Semantics

Dr. David Koop
Reminders

- Project Emails:
  - Provide updated project proposal
  - Confirm that I understand what your project is about

- Project Progress Reports Due November 13
Semantics

• “The meanings of words and phrases in a particular context” [Merriam-Webster Dictionary]

• Computers understand what they are programmed to understand
  - In general, these are very structured tasks that require only specifically programmed knowledge
  - How could computers understand a subject in the same way we do?
    • Natural language processing
    • The knowledge itself (concepts, definitions, etc.)
    • How data items relate
Introduction to Semantic Web Technologies & Linked Data

Oktie Hassanzadeh
Terms

• Ontologies: formal frameworks for representing data
  - OWL is a language to author ontologies
  - Often curated, not as much reuse

• Semantic Web / Linked (Open) Data / Web of Data
  - Markup data with machine-interpretable information
  - Data represented using RDF (Resource Description Framework)
  - RDF is **triples** (subject-predicate-object)
  - Can be “dirtier”, less well-specified, multiple definitions

• SPARQL: **SPARQL Protocol and RDF Query Language**
  - Way to query linked data
  - Support recursion
Janus: from Workflows to Semantic Provenance and Linked Open Data

P. Missier, S. S. Sahoo, J. Zhao, C. Goble, and A. Sheth

Presented by: Yashwanth Nandanam
Linked Open Provenance

P. Missier

http://www.slideshare.net/pmissier/nesc-invited-presentation-semantic-provenance-and-linked-open-data
Janus

- Provenir: ontology for describing workflow and provenance concepts
- Domain-specific ontologies: domain-specific information (e.g. biology terms, etc.)
- SQUIN: SPARQL queries over multiple linked data sources
  - More efficient means of querying provenance with domain-specific information
- Tie different identities together: identifiers from the provenance world and identifiers from domain-specific ontologies
Provenance and Semantics

• What about provenance for linked data itself?
  - What if someone publishes data that is incorrect?
  - e.g. “Toronto is a US City”

• Using Provenance in the Semantic Web (Web Semantics Journal Special Issue)
  - DBPedia work [F. Orlandi and A. Passant]
  - Reasoning including trust annotations [Bonatti et al.]
  - Link with reproducibility [L. Moreau]
  - Workflow tools [E. Pignotti et al.]
Reminders

• Project Emails:
  - Provide updated project proposal
  - Confirm that I understand what your project is about

• Next Class:
  - Provenance Standards (OPM, PROV)
  - Reading is a Web document