Choosing between Axioms, Rules & Queries: Experiments with Semantic Integration Techniques

Christopher J. Matheus, Bell Labs Ireland
OWLED – June 6, 2011
Semantic Data Access @ BLI

• Semantic Data Access
  – Research effort based at Bell Labs Ireland
  – Staffing: 3 MTS @ BLI, 2 MTS @ BL US, 2-6 new hires over 4 years, 4 PhD students
  – Collaborations: DERI/NUI, FAME/TCD, CLARITY/UCD, Oxford
  – Mission: to research the use of semantic technologies to ease the access to large distributed heterogeneous data sources/services within the general telecommunications domain
  – Approach: investigate state of the art, conduct empirical experiments, develop functional prototypes
Fundamental Research Questions

• FQ1: Representation
  What are the appropriate conceptualizations (i.e., ontologies) for the domains of interest to ALU customers?

• FQ2: Lifting/Alignment
  How can distributed, heterogeneous data be effectively lifted/aligned into a useful semantic layer of abstraction?

• FQ3: Reasoning
  How do we efficiently reason (infer new facts) about data using various semantic techniques?

• FQ4: Service Description
  How can services be semantically described so as to facilitate their discovery and composition?
Three Integration Techniques

• OWL axioms
• SWRL rules
• SPARQL queries

• Question: can we experimentally assess the relative costs and benefits of each technique under different problem domains and characteristics in terms of performance (space and time) and ease of use?

• Ultimate goal: develop design patterns and best practices to guide developers in the appropriate use of semantic technologies
Three Problem Domains

• Smart Conference/Campus
  – Integrating FOAF, DBLP and simple location data

• Femto Cell Network (BLI Testbed)
  – Integrating network management, location, and contact/social data

• Wireless Sensor Networks
  – Integrating network management, location, contact/social, and sensor data
Smart Conference Experiments*

• Conference Scenario: identify and locate attendee’s “acquaintances”

• An acquaintance is some one in an attendee’s FOAF file or someone they co-authored a paper with in DBLP

• Problem: define and evaluate this acquaintance relationship using the three integration techniques

*J. Keeney, A. Boran, I. Bedini, C.J. Matheus, P.F. Patel-Schneider, Approaches to Relating and Integrating Semantic Data from Heterogeneous Sources, 2011 IEEE/WIC/ACM International Conference on Web Intelligence.
OWL Definition

• OWL axioms:
  – SymmetricProperty(sda:acquaintance)
  – SubPropertyOf(foaf:knows sda:acquaintance)
  – SubPropertyOf(
    owl:ObjectPropertyChain(
      foaf:maker ObjectInverseO(foaf:maker))
    sda:aquaintance).
SWRL Definition

• FoafRule:
   (?Person2 foaf:knows ?Person1) ->
   (?Person1 sda:acquaintance ?Person2)
   (?Person2 sda:acquaintance ?Person1)

• AuthorRule:
   (?Document foaf:maker ?Person1)
   (?Document foaf:maker ?Person2) ->
   (?Person1 sda:acquaintance ?Person2)
   (?Person2 sda:acquaintance ?Person1)

• SameAs1: (?x owl:sameAs ?y) (?x ?p ?o) -> (?y ?p ?o)
• SameAs2: (?x owl:sameAs ?y) (?s ?p ?x) -> (?s ?p ?y)
SPARQL Definition

• WHERE {
  ?Person1 foaf:name "John Doe".
  ?friend foaf:name ?friendname
}

CONSTRUCT {
  ?Person1 foaf:name "John Doe".}

High-level Results

• SPARQL: best performance (size, max-size, time) but difficult to write correctly
• SWRL: even harder to write/read
• OWL: easiest to write by far but not able to handle largest datasets
• Caveats:
  – initial pass at this type of experiment
  – one simple problem
  – different reasoners: Pellet for OWL, Jena for SWRL, Jena TDB for SPARQL
  – subjective measure of definition complexity