Ontology Summit 2014 Track B: Making Use of Ontologies – Tools, Services, Techniques Synthesis (1st round)

Alan Rector, Christoph Lange

20 Feb 2014

Mission Statement in a Nutshell

The Web of Data ...

- provides great opportunities for ontology-based services, but also
- puts challenges to
 - tools for editing and using ontologies, and to
 - techniques for ontological reasoning and ontology engineering.

Terminology: Web of Data, Big Data, Semantic Web

We use "Web of Data" to subsume both ...

- Big Data (w.r.t. volume, velocity, variety)
- Semantic <u>Web</u>
 - making sense of knowledge distributed over the Web
 - not just using IRIs as local symbol names

Issues and Topics for Comments

- Relevance of Ontologies to Big Data
 - Building vs mining vs induce vs direct queries to large data stores
 - Do even light-weight ontologies scale?
 - What is realistic for ontological commitments for big heterogeneous data?
 - Ontologies for annotating big data vs ontologies for representing big data
 - Bio community has massive annotation effort dependent on Gene Ontology and related.
- Variety/heterogeneity and re-use
 - Almost all big data is heterogeneous (→ Track D)
 - How do tools relate to re-use (→ Track A)
- Is all Knowledge Representation "ontology"?
 - Should we factor the universal from particular axioms from templates from rules?
- Relation amongst formalisms and relations to traditional IT
 - OWL, RDF, XML, UML, CGs, CL, Rule languages, ...
 - Interaction of open world & closed world assumptions
- Limitations of tools (and of formalisms) Fit tools to problems rather than problems to tools!
- "Webification" of pre-Web formalisms Retrofitting IRI/URI references and Linked Data conformance
- Requirements? We can't even say what we need

Relevance of Ontologies to Big Data: Potential

- IBM Watson (ChrisWelty) a service that
 - Answers rich natural language questions over a broad domain of knowledge
 - ... giving precise answers with an accurate assessment of confidence and consumable justifications – within seconds
- Reasoning Approach:
 - Don't build a formal ontology of the World (and unify formal logical representations of the questions with it),
 - but locally learn ontologies on demand, drawing on formal as well as informal sources, using different reasoning techniques:
 - 1. Generate hypotheses
 - Evidence retrieval
 - Includes keyword matching against as-is natural language text sources
 - Challenge: disambiguating types of entities and predicates (partly solved using existing taxonomies)
 - 3. Evidence scoring: largely based on machine learning (i.e. statistical techniques)

Relevance of Ontologies to Big Data: Challenges

- Web-wide data unlikely to conform to a single tight ontology
 - Watson limits itself to a few simple taxonomies (e.g. YAGO) for entity/property typing as part of Entity Disambiguation & Matching(EDM)
 - Some large collaborations may agree on a limited subset of ontolologies
 - e.g. Parts of molecular biology and the Gene Ontology and other OBO Foundry ontologies
 - (Lack of convincing use of ontologies in big data examples in forum presentations so far)
- How to create ontologies from data
 - Build labour intensive
 - Mine / re-use inconsistencies, incompleteness, irrelevance of data "out there"
 - Machine learning is it ready?
 - What combination for what problems? (Watson is making first steps beyond Jeopardy: In health care)
- What for? Ontologies for representation vs ontologies for annotation & indexing
 - Many bio examples of big data annotated with terms ontologies.

Variety & Heterogeneity

- Of Information types
- Of schemas
- Of software
 - OntolOp standardization effort (TillMossakowski):
 - Distributed Ontology Language: meta language for ontology and data languages
 - Hets and Ontohub tools support alignment and reasoning across ontology languages
- Tools not currently ready for "big w.r.t. volume", but OntolOp standard paves the way:
 - Splitting big ontologies into modules
 - Distributing interlinked modules over the Web
 - Linked Data conformance retrofitted into pre-Web ontology languages

Is all information ontology?

- What there is, or everything we know about what there is?
 - AlanRector: Universal/Essential vs
 Contingent/Accidental/Particular (more realistic on the Web!)
 - Should we have the same representation for all?
 - Should our architectures partition the knowledge formally?
- Interaction with languages
 - OWL/DLs good for open world universal knowledge
 - Template formalisms frames, UML, rules good for contingent knowledge

Relations amongst Formalisms not well understood

- OWL/Description logics, other logics
 - Intensional descriptions with model theoretic semantics
 - Most semantics is in the language itself
 - SPARQL OWL entailment regime still very new & little experience or tools
- RDF/RDF(S) & SPARQL
 - Practical usage much greater OWL/(DavidPrice),
 - Basis of many big data systems: Copes (fairly) well with heterogeneous data
 - Most of the semantics in the SPARQL queries; minimal semantics in language itself
- UML & relational database schemas
 - Widely used and understood
 - Good visualisation
 - Interaction of UML and OWL a critical problem for some users
 - UML wasn't originally intended to be formalized
 - OntolOp has so far adopted one out of many possible formal semantics for UML
 - Relates to relation of Ontology's to Knowledge Representation
- Rules (See separate rule stream in Ontolog seminars 2013)
 - Not well standardised but a good fit for many problems

Limitations of Tools

- Most Inherited from hand building of small ontologies for specific applications
 - Most limited to a single, or a few, formalisms (similar for techniques)
 - "Silo-ing"
- Approaches for breaking the tool boundaries:
 - One can take inspiration from template formalisms in using OWL (AlanRector)
 - OWL is widely supported by tools that work, which makes it attractive and usable (AndreaWesterinen)
 - OntoIOp related tools integrate many universal-knowledge languages, but also first steps towards frames, UML, rules (TIIIMossakowski)
- Visualisation a major problem
- Few proven to scale
 - Scaling of reasoners/theorem provers known to be problematic, although orders of magnitude improved in past 5 years.

Requirements

(Editorial Comment)

- Nobody can say what they want
 - Until somebody shows it, people don't know they need it
 - Few human factors studies
 - (Manchester has one it would be happy to present)
 - Users wanted to do a wider variety of tasks altogether than any one set of tools supported – semantic, syntactic, lexical, linguistic, web searching,
 - Needs reflection on exemplars
 - What would it have taken to build this more efficiently and effectively?