Challenges in Scaling Tools for Ontologies to the Semantic Web: Experiences with Hets and OntoHub

Till Mossakowski

Otto-von-Guericke University of Magdeburg

Ontology summit 2014
Track B: Making use of Ontologies: Tools, Services, and Techniques
Ontologies for the Semantic Web
The Semantic Web Layer Cake

User Interface & Applications

Trust

Proof

Unifying Logic

Ontology: OWL

Rule: RIF

Query: SPARQL

RDFS

Data interchange: RDF

XML

Unicode/IRI
Tim Berners Lee’s early semantic web layer cake

Semantic Web = unifying language for classical logic = bus

- Conversational interfaces
- Heuristics e.g. search engines
- Non-classical "logics"

- Trusted system e.g. access control
- Digital signature
- Bank machine
- KR inference engine
- Haystack inference engine

- Higher order
- First order
- Propositional Logic

RDF

XML
Challenges

- Diversity of data and ontology languages:
  - RDF, RDFS, OWL, RIF, SparQL, FOL, HOL
  - actually even more: EER, OBO, UML, SKOS, F-logic, Common Logic, SBVR, …

- matching and alignment — also across languages
- merging/combination along alignments
- distributed data and ontologies
- ontology-based data access/management
OntoIOP
Need for a Unifying Meta Language

Not yet another ontology language, but a meta language covering

- diversity of ontology and data languages
- translations between these
- diversity of operations on and relations among ontologies

Current standards like the OWL API or the Aligment API only cover parts of this

The Ontology, Modeling and Specification Integration and Interoperability (OntoIOp) initiative addresses this

http://ontoiop.org
The OntoOp initiative

- started in 2011 as ISO 17347 within ISO/TC 37/SC 3
- now continued as OMG standard
  - request for proposals (RFP) has been issued in December 2013
  - proposals answering RFP due in December 2014
- 50 experts participate, ~ 15 have contributed
- OntoOp is open for your ideas, so join us!
- Distributed ontology language (DOL) is being prepared as one answer to RFP
An Initial Logic Graph for OntoIOp
Tools for OntoOp/DOL
Heterogeneous Tool Set (Hets, hets.dfki.de)

Architecture of the heterogeneous tool set Hets

Tools for specific logics

- Text Parser
- Abstract syntax
- Static Analysis (Signature, Sentences)
- Interfaces
  - XML, Aterm

Theorem provers
Rewriters

Conservativity checkers
Model finders
Model checkers

Logic graph

- Haskell ...
- Isabelle
- LF
- HasCASL
- CspCASL
- CoCASL
- VSE
- SoftFOL
- TPTP
- Maude
- ModalCASL
- QBF
- RelScheme
- CommonLogic
- Propositional

Grothendieck logic
(Flattened logic graph)

Tools for heterogeneous specifications

- Text Parser
- Abstract syntax
- Static Analysis
- Global Environment
- Interfaces
  - XML, Aterm
  - WWW, GUI

- Heterogeneous development graphs
- Decomposition of proof obligations
- Management of proofs
- Heterogeneous inference engine
- Heterogeneous proof trees

Mossakowski
Challenges in Scaling Tools: Experiences with Hets and OntoHub
2014-01-30 11
Ontohub web portal and repository

Ontohub is a web-based repository engine for distributed heterogeneous (multi-language) ontologies

- prototype available at http://ontohub.org
- speaks DOL, RDF, OWL, Common Logic, MOF, and other languages
- git repositories as version control backend
- Hets as reasoning backend
- annual Ontology Summit as a venue for review, and discussion
Semantic web challenges for Ontohub and Hets

- move from identifiers to IRIs and CURIEs
- linked data compliance
- large amounts of data
Move from identifiers to IRIs and CURIEs

- Hets and Ontohub now use IRIs everywhere
  - problem: the involved languages do not necessarily do so
  - solution: interpret identifiers relative to based URIs

- problem: there is no well-established module system for IRIs
  (OWL only provides imports, no complex modules)
- solution: use MMT URIs or form doc?mod or doc?mod?sym
  (doc: document URI, mod: module name, sym: symbol name)
  - MMT: combines a module system, a foundationally uncommitted formal semantics, and web-scalable implementations

- CURIEs are a useful abbreviation mechanism, taken from RDFa
Linked data compliance

- Ontohub is linked-data compliant
  - at ontohub.org/colore/algebra/vector.clif, you can download a CLIF specification of vector spaces in various formats (HTML, CLIF, XML)
- practically no other ontology repository is linked-data compliant
  - URL catalog mechanisms help to remedy this problem
  - need to mirror external ontologies and repositories in Ontohub
Large amounts of data

- modularity helps
  - Hets and Ontohub support a rich module language: DOL
  - Hets and Ontohub can display module graphs
  - module graphs are split into sub-graphs (one for each document), with external references
- DOL supports module extraction
- re-use of logical theories, search for logical commonalities
- service decomposition, server farm for parsing, analysis and proving
- BUT: no big data (Ontohub database is PostgreSQL)
Two module graphs, too large
A module graph, handy

Ontologies for the Semantic Web

OntoloOp

Tools for OntoOp/DOL

A module graph, handy

Mossakowski Challenges in Scaling Tools: Experiences with Hets and OntoHub 2014-01-30 18
Future work

- tools for splitting ontologies
- tools for selecting relevant axioms
- noSQL databases for ontology repositories?