Wither OWL in a knowledge-graphed, Linked-Data World?

Jim Hendler
@jahendler

Tetherless World Professor of Computer, Web and Cognitive Science Director, Rensselaer Institute for Data Exploration and Applications Rensselaer Polytechnic Institute
Semantic Web
Where are we now?

Now

You are here!

Later

Slide ca. 2001. Were we right?
Were we right?
Sort of?  Mostly?
Tetherless World Constellation
Example: Semantic Search

Web 3.0: The Dawn of Semantic Search

James Hendler, Rensselaer Polytechnic Institute

Emerging Web 3.0 applications use semantic technologies to augment the underlying Web system's functionalities.

IEEE Computer, Jan 2010
Contenders ca. 2010

Tetherless World Constellation

List compiled and created by Kevin Eklund @ ToMuse.com
Google Sem Webbers include: R. Guha, Dan Brickley, Denny Vrandečić, Natasha Noy, Chris Welty, …
Google finds embedded metadata on >20% of its crawl – Guha, 2014
Other success stories

Facebook: 2011
Oracle: 2012
It's not so hard to find out, there are already lots of companies using SW technologies, some of which are:

- New York Times (data.nytimes.com)
- Facebook (The Open Graph protocol)
- Google (The Knowledge Graph)
- IBM (Watson, see also this interview)
- BBC (World Cup 2010)
- Boeing (Testimonials)
- O'Reilly (Semantic SEO)
- Rotten Tomatoes (Toy Story 3)
- Monster.com (Semantic Search)
- Adobe (XMP)
- eBay (Headphones)

Here is a recent article about Semantic Web technologies at BBC: http://www.cmswire.com/cms/information-management/bbcs-adoption-of-semantic-web-technologies-an-interview-017981.php
• Which of these use OWL in any significant way?

(This part of the slide intentionally left blank)
• This is NOT to say OWL isn’t being used
  – but it’s not much by comparison
  – but it’s not much on the Web
    • with the possible exception of the misuse of owl:sameAs, but let’s not go there...

• Semantic markup on the Web exceeds anything we predicted in 2001...
  – ... but OWL use on the Web as a proportion of this lags behind the expectations many of us had
    • The rest of this talk is speculation as to why...
ROI: Web 3.0

• The "small o" ontology finds use cases in Web Applications (at Web scales)
  – A lot of data, a little semantics
  – Finding anything in the mess can be a win!

• Example
  – Declare simple inferable relationships and apply, at scale, to large, heterogeneous data collections
    • These are "heuristics" not every answer must be right (qua Google)
      – But remember time = money!
WHAT I NOW BELIEVE

ROI: Web 3.0

• The "small o" ontology finds use cases in Web Applications (at Web scales)
  – A lot of data, a little semantics
  – Finding anything in the mess can be a win!

• Example
  – Declare simple inferable relationships and apply, at scale, to large, heterogeneous data collections
    • These are "heuristics" not every answer must be right (qua Google)
      – But remember time = money!
• That explains why the uptake of RDFS/SPARQL is going on
  – and used by things like schema.org
• But still doesn’t explain why OWL isn’t used as much as it should be
  – especially on the Web
  – especially when the need for ontologies is growing rapidly and many kinds of “ontology like things” are being used heavily
• My previous view was blaming OWL’s problems on too much expressivity
  – as opposed to RDFS (or really RDFS+)
• I now believe the problem is lack of expressivity (in an interoperable way)
  – for tasks where people really need Web semantics
The Web is increasingly about LINKING DATA

http://linkeddata.org/ cloud is a tiny fraction of what is out there
Current linked data approaches miss a key need

- Data is converted into RDF and SPARQL’ed
  - creates huge graph DBs less efficient than the original DB
- Data is converted from DB into SPARQL return on demand
  - much better, but you must know the mapping
- owl:sameAs is (ab)used to map data to data
  - but that only lets you map equals – which is an easy mapping to express in many ways
    - defining equality right in a model theory is much harder, and thus the abuse, but let’s leave that for another talk
Linked Semantic Web “metadata” documents that can be used to link very large databases in distributed data systems. This could lead to orders of magnitude reduction in information flow for large-scale distributed data problems.
• We want a knowledge representation that can do things like:
  – help us find the right data for a problem
    • The “Date” field in some DB could be lots of different things
      – consider “Database of 1957 NYC births” vs “Database of 1957 NYC deaths”
  – help us map between different databases
    • The problem isn’t primarily translating ontologies to each other, it is tying the ontologies to the data (for the mappings)
• OWL doesn’t have a part-whole relation
  - left out of design because we couldn’t reach consensus (and there’s 2000 yrs of argument behind that)
  - but also because most require transitive closure of parts in many cases and that had complexity issues
  - but one of the most used relations in the gene ontology and many medical ontologies is part of
Note in the above – we use the fact that the brain has 5 lobes as a driving example for qualified cardinality - but to say that in OWL you need to INVENT the “hasdirectpart” relation.
• Many things in database schemas also map to parts/wholes
  – Who is in what organization?
  – What components comprise an assembly?
  – Where did something occur (that was part of another event)?
  – When...
• Temporal reasoning is missing from OWL
  – Knowing [this talk occurred during “NSF-MTG 2016” AND “NSF-MTG 2016” occurred in May of 2016, THEREFORE this talk occurred in May of 2016] is important
  • Whole books of temporal logics out there
    – picking is hard
    – but it is also what OWL has to do to be a standard for this kind of mapping
• Add math to part-wholes
  – OY!
  – but GDBs are widely used for mappings of information in big databases
    • especially in science (more OWL use cases)

• Talking about “math”
  – lots of discussion of adding probability to OWL
    • and someone should some day
  – but what about
• Discovery Informatics and data mining
  – Huge industries, and growing
  – Web data, Internet of things, data interoperability (the startup holy grails)
• Example: supposing some scientific theory “implies” that if X increases then Y should also increase
  – Which databases would help confirm my theory? Which would argue against it?
  – Easy to check in many new database systems
  – How do we express that aspect of the theory in OWL?
• In fact, what about procedural attachment?
  – lots of literature on preserving completeness w/respect to procedures
    • used in prolog, etc.
  – Why isn’t this in OWL?
    • patent issues w/respect to standard
    • general concern about whether procedural attachment was possible in OWL DL framework
No longer a patent issue …
• There are many examples
  – Describing how data in one set could be joined to data in another is incredibly powerful, timely, and important
    • it’s just not really what people have typically used traditional reasoners for
  – Providing the ontological glue among different AI technologies: Priceless

• Why aren’t we doing more to understand these issues and bring into the OWL “family?”
  – de facto standardization leads to adoption
• Is the problem that we cannot have these powerful things in a decidable (sound and complete) language?
  – then maybe we have to give up decidability?
• or even better, define new maths of expressivity that have different kinds of “sound and complete” behavior
  – i.e. approximately sound and complete
    » (is modern computational theory weakened by “within epsilon” optimality?)
  – i.e. “anytime” algorithms for reasoners
    » (conjecture) sound and complete at infinity
• We may need to rethink how we are defining ontologies with respect to the necessary properties for use on the Web
I’m not “anti-formalism”

• A sufficient formalism for Semantic Web applications must
  – Provide a model that accounts for linked data
    • What is the equivalent of a DB calculus?
  – Provide a means for evaluating different kinds of completeness for reasoners
    • In practice we must be able to model A-box effects as formally as T-box technologies
      – example Weaver 2012 showed what restrictions were needed to maximize parallelism of some OWL subsets
  – Think about other processors than formal reasoners that will use the ontologies
    • ontologies used in many other ways
      – i.e. why does an IE system w/F1=.84 need a DL reasoner?
The growing world of (semantic web, linked) data needs ontologies more than ever
  - OWL has some of the important things
  - But is missing many of the really important things

The problem isn’t (formal) expressivity
  - it’s the need to express other things (esp relationships between properties, events, etc.)
  - We need more research into how to formalize these kinds of relations
I get funded by lots of folks – this talk may or may not represent anything anyone there believes:
- ARL, DARPA, Microsoft, Mitre, Optum Laboratories

The Rensselaer Institute for Data Exploration and Application pays my salary
- Thanks!!

Many of my best ideas, including a lot in this talk come from listening to smart people
- Frank van Harmelen and Peter Mika have influenced my thinking about many of the ideas in this talk

The students and my colleagues at the RPI Tetherless World Constellation (tw.rpi.edu)
knowledge representation

facts → reasoning engine → answers