

# **Semantic Web**

## outlook and trends

# The Past ~30 Odd Years

- 1984 Lenat's Cyc vision
- 1989 TBL's Web vision
- 1991 DARPA Knowledge Sharing Effort
- 1996 RDF
- 1998 XML
- 1999 RDFS
- 2000 DARPA Agent Markup Language, OIL
- 2001 W3C Semantic Web Activity
- 2003 OWL
- 2008 SPARQL
- 2009 OWL 2
- ~2009 Linked Data
- 2011 Schema.org
- 2012 Wikidata
- 2012 Microdata & schema.org
- 2013 Rule Inter. Format
- 2009- vocabularies: SKOS, PROV, RDB2RDF, ...
- 2014 JSON-LD
- 2017 SHACL

**The Next 30?**

**?**

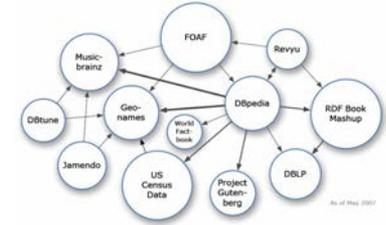
# What's Hot



Here are six areas that I think will be important in the next five years

- Linked Data
- Semantic Data
- Big (Semantic) Data
- Populating RDF KGs from text
- Schema.org
- Wikidata
- Machine learning and structured data

# Linked Data



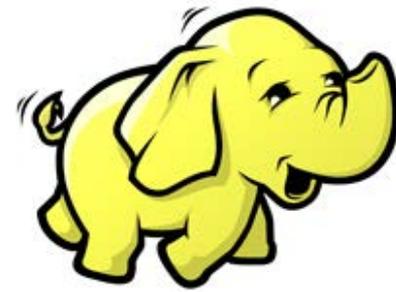
- RDF is a good data language for many applications
  - Schema last applications, graph model is easy to map into others, Web oriented
- OWL is a poor KR language in many ways
  - no certainties, contexts, default reasoning, procedural attachments, etc. Current OWL most rely on forward reasoning and don't handle contradictions well.
- Today's immediate benefits mostly come from shallow reasoning and integrating and exploiting data rather than reasoning with deeper “ontological knowledge”

# “Semantic” Data



- The S word is very popular now
- Semantic  $\neq$  Semantic Web
- Search companies are competing by better understanding (i) content on a web page and (ii) a user’s query
- Facebook benefits from its social graph: you say you attended UMBC, not “UMBC”. FB knows it’s a university, which is a kind of educational institution
- Hendler: “A little semantics goes a long way  
– It’s incremental: don’t try to do it all at once

# Big (Semantic) Data



- The big data theme and the growth of RDF data combine to create a need for better semantic tools that can work at Web scale
- Problems include:
  - Parallel reasoning (Hard, see [Webpie](#) paper & letters)
  - Distributed SPARQL queries
  - Graph analytics on huge RDF graphs
  - Machine learning over RDF data
  - Extracting and using statistical knowledge from RDF

# Knowledge Base Population



- Information extraction involves extracting entities and relations from text
- A common model: read lots of text documents and populate a knowledge Base with the entities, attributes and relations discovered
  - See DARPA Machine Reading Program, NIST TAC Knowledge Base Population track
- RDF/OWL is increasingly chosen as the default target for such knowledge

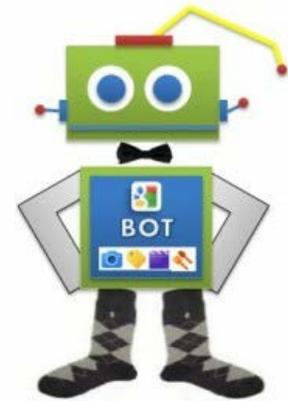
# TAC 2012

# Cold Start

Knowledge Base Population



# Microdata aka Schema.org



- It's significant that the big search companies have embraced an RDF compatible way to embed data in Web pages
- They are beginning to detect and exploit the data to provide better services
- It demonstrates that it's not rocket surgery, is easy to add, and is useful
- Their measured incremental approach is pragmatic and will open up possibilities for more

# Wikidata



- Wikipedia has been enormously successful and important, making all of us smarter
- DBpedia shows its potential to make machines more intelligent
- Wikidata aims to better integrate these two and has the potential of creating a knowledge resource with a permeable barrier between the unstructured and structured representations

# New Application Areas

Some application areas will get a lot of attention because they are important or new

- Cybersecurity: Modeling cyber threat intelligence
- Healthcare: Electronic healthcare records, personalized medicine
- Mobile computing: Modeling and using context, integrating information from phone, web, email, calendar, GPS, sensors, etc.
- Ecommerce: E.g., GoodRelations

# Beyond PDF

- Publication is important to all scholarly disciplines, especially STEM areas
- Modernizing this is more than putting pdf versions of articles online
- There is an interest in also publishing data, services and code and linking these to papers
  - Capturing provenance is an interesting aspect
  - [Google Dataset Search](#)
- We need new author tools, indexing services, search engines, etc.

# Conclusion

- We are still exploring what can be done
  - and how to do it
  - and how to do it efficiently
  - and how to do it easily w/o a lot of training
  - and how to derive benefits from it (commercial or societal)
- The technology and systems will change
- It will be a fluid area for another decade or two
  - or maybe longer