Knowledge Graphs
outlook and trends

Dec 2022

The Past ~40 Years

- 1984 Lenat's Cyc vision
- 1989 TBL's Web vision
- 1991 DARPA Knowledge Sharing
- 1996 RDF
- 1998 XML
- 1999 RDFS
- 2000 DARPA DAML, OIL
- 2001 W3C Semantic Web push
- 2003 OWL
- 2007 Property graphs (neo4j)
- 2008 SPARQL
- 2009 OWL 2

- 2009 Linked Data
- 2010s Commercial systems
- 2011 Schema.org
- 2011 KG embeddings
- 2012 Wikidata
- 2012 Microdata & schema.org
- 2013 Rule Interchange Format
- 2009- ontologies: SKOS, PROV, …
- 2014 JSON-LD
- 2017 SHACL, ShEx
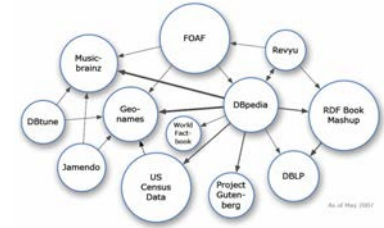- 2019 GQL
- 2022 RDF*
- …

The Next 5 or 10?

?

What's Hot

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

- Big (Linked)(Semantic) Data
- Community KGs like Wikidata
- KGs and natural language processing (NLP)
- Embedded KG data, e.g., Schema.org
- Machine learning and KGs
- Integrating KGs with large language models
- Applications, including property graphs

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"

Community KGs

- 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 unstructured and structured representations

KGs ⬄ NLP

- NLP information extraction can be used to extract data and populate a KG
  - Keep specialized KGs up to date, e.g., adding information about new Cybersecurity threats
- KGs can disambiguate and link NLP mentions
  - E.g., which place named Springfield is this?
  - Wikidata finds 53 human settlements named Springfield in the US. Using text around the mention and the properties can help identify the right one

Microdata aka Schema.org

- Big search companies embraced an RDF compatible way to embed data in Web pages

- Increasingly they detect and exploit the data to provide better services

- Schema.org's ontology is growing to cover more concepts and properties

- The measured incremental approach is pragmatic and has opened up possibilities

"Semantic" Data

- The S word is very popular now
- Semantic ≠ 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
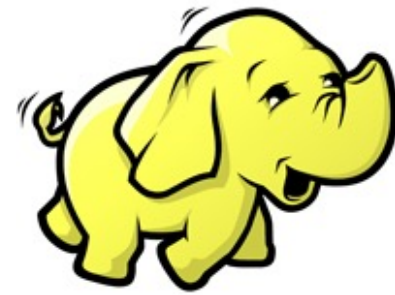
- The big data theme and the growth of RDF data combine to create a need for better semantic tools that can work at Web scale

- Problems include:
  - Parallel reasoning (Hard, see Webpie paper & letters)
  - Distributed SPARQL queries
  - Graph analytics on huge RDF graphs
  - Machine learning over RDF data, graph embeddings
  - Extracting and using statistical knowledge from RDF

New Application Areas

Some application areas will get a lot of attention because they 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
- **Science:** specialized KGs for disciplines like Biology

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 at least another decade