Introduction to the Semantic Web

Questions

• What is the Semantic Web?
• Why do we want it?
• How will we do it?
• Who will do it?
• When will it be done?

“XML is Lisp's bastard nephew, with uglier syntax and no semantics. Yet XML is poised to enable the creation of a Web of data that dwarfs anything since the Library at Alexandria.”


“The web has made people smarter. We need to understand how to use it to make machines smarter, too.”

-- Michael I. Jordan, paraphrased from a talk at AAAI, July 2002 by Michael Jordan (UC Berkeley)
“The Semantic Web will globalize KR, just as the WWW globalize hypertext”

-- Tim Berners-Lee

“The multi-agent systems paradigm and the web both emerged around 1990. One has succeeded beyond imagination and the other has not yet made it out of the lab.”

-- Anonymous, 2001

IOHO

- The web is like a universal acid, eating through and consuming everything it touches.
  - Web principles and technologies are equally good for wireless/pervasive computing
- The semantic web is our first serious attempt to provide semantics for XML sublanguages
- It will provide mechanisms for people and machines (agents, programs, web services) to come together.
  - In all kinds of networked environments: wired, wireless, ad hoc, wearable, etc.

Origins

Tim Berners-Lee’s original 1989 WWW proposal described a web of relationships among named objects unifying many information management tasks.

Capsule history

- Guha’s MCF (~94)
- XML+MCF=>RDF (~96)
- RDF+OO=>RDFS (~99)
- RDFS+KR=>DAML+OIL (00)
- W3C’s SW activity (01)
- W3C’s OWL (03)

http://www.w3.org/History/1989/proposal.html
**W3C’s Semantic Web Goals**

**Focus on machine consumption:**

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."


---

**TBL’s semantic web vision**

![Semantic web stack 2006 diagram]

**Why is this hard?**

---

**Semantic web stack 2006**

![Semantic web stack 2006 diagram]

---

**Why is this hard?**

after Frank van Harmelen and Jim Hendler
What a web page looks like to a machine...

after Frank van Harmelen and Jim Hendler

OK, so HTML is not helpful

Maybe we can tell the machine what the different parts of the text represent?

title
speaker
time
location
abstract
biosketch
host

XML to the rescue?

XML fans propose creating a XML tag set to use for each application.

For talks, we can choose <title>, <speaker>, etc.

after Frank van Harmelen and Jim Hendler

XML ≠ machine accessible meaning

But, to your machine, the tags still look like this....

The tag names carry no meaning.

XML DTDs and Schemas have little or no semantics.

after Frank van Harmelen and Jim Hendler
XML Schema helps

XML Schemas provide a simple mechanism to define shared vocabularies.

But there are many schemas

There’s no way to relate schemas.

An Ontology level is needed

We need a way to define ontologies in XML
So we can relate them
So machines can understand (to some degree) their meaning.
Semantic web technologies allow machines to share data and knowledge using common web language and protocols.

~ 1997

Semantic Web beginning

Data is inter-linked to support integration and fusion of knowledge

2007

LOD beginning

Use Semantic Web Technology to publish shared data & knowledge

Semantic Web => Linked Open Data

Data is inter-linked to support integration and fusion of knowledge

2008

LOD growing

Use Semantic Web Technology to publish shared data & knowledge

Semantically Linked Open Data

Data is inter-linked to support integration and fusion of knowledge

2009

... and growing

Use Semantic Web Technology to publish shared data & knowledge

Semantically Linked Open Data

Data is inter-linked to support integration and fusion of knowledge

... and growing
**Linked Open Data**

Use Semantic Web Technology to publish shared data & knowledge

LOD is the new Cyc: a common source of background knowledge

Data is interlinked to support integration and fusion of knowledge

2010...growing faster

**Today and tomorrow**

- Simple ontologies like FOAF & DC in use today
  - We’ve crawled more than 3M FOAF RDF files
- We hope to be able to make effective use of ontologies like Cyc in the coming decade
  - There are skeptics...
  - It’s a great research topic...
- The SW community has a roadmap and some experimental languages...
- Industry is still holding back...
  - They are being conservative
- We need more experimentation and exploration

2011: 31B facts in 295 datasets interlinked by 504M assertions on ckan.net