

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.”

-- Philip Wadler, *Et tu XML? The fall of the relational empire*, VLDB, Rome, September 2001.

“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.

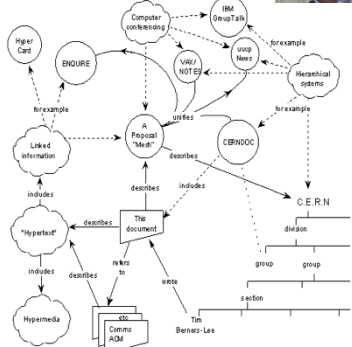
Origins

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

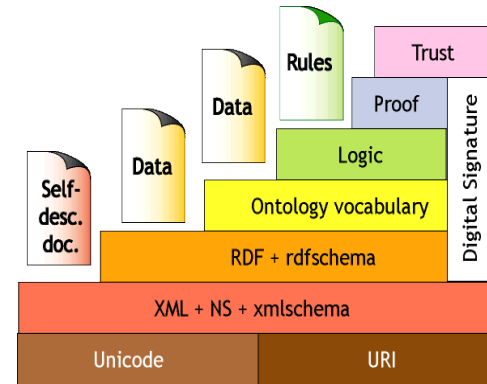
W3C's Semantic Web Goals

Focus on machine consumption:

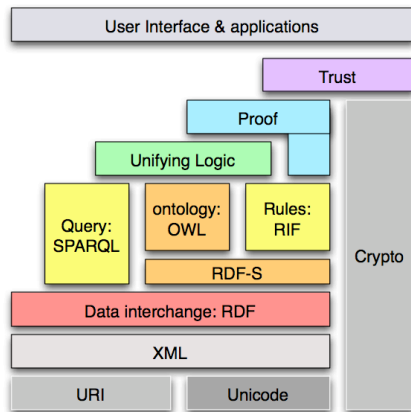
"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

-- Berners-Lee, Hendler and Lassila, The Semantic Web, Scientific American, 2001

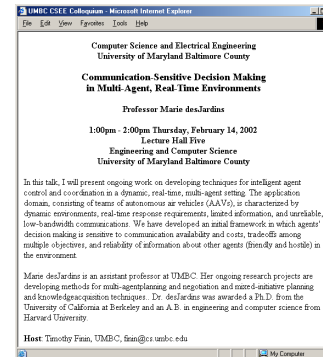
TBL's semantic web vision



Semantic web stack 2006

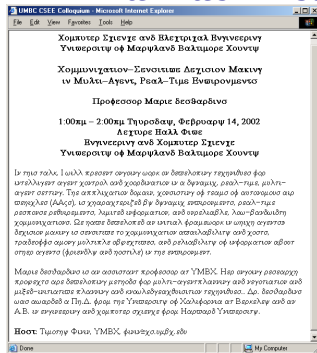


Why is this hard?



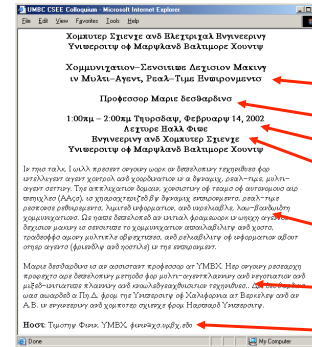
after Frank van Harmelen and Jim Hendler

What a web page looks like to a machine...



after Frank van Harmelen and Jim Hendler

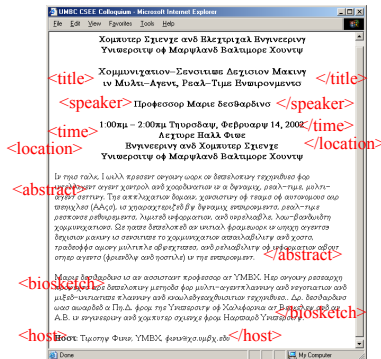
OK, so HTML is not helpful



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?

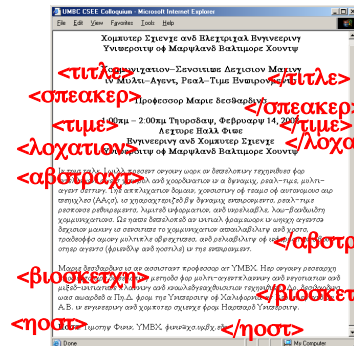


XML fans propose creating a XML tag set to use for each application.

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 Schema file

XML Schemas provide a simple mechanism to define shared vocabularies.

XML Schema file 1

XML Schema file 42

Annotations: <title>, </title>, <abstract>, </abstract>, <location>, </location>, <abstract>, </abstract>, <biosketch>, </biosketch>, <host>, </host>

after Frank van Harmelen and Jim Hendler

But there are many schemas

XML Schema file 1

XML Schema file 42

XML Schema file 1

XML Schema file 42

Annotations: <title>, </title>, <abstract>, </abstract>, <location>, </location>, <abstract>, </abstract>, <biosketch>, </biosketch>, <host>, </host>

after Frank van Harmelen and Jim Hendler

There's no way to relate schema

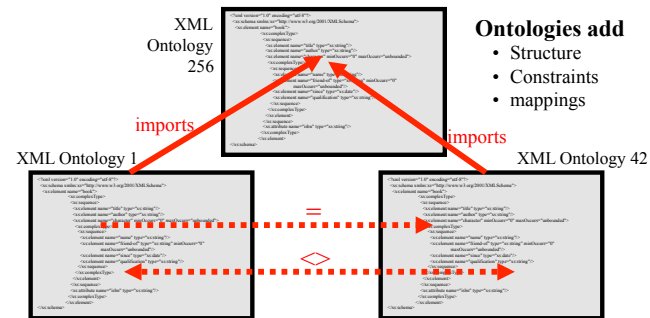
XML Schema file 1

XML Schema file 42

Annotations: <title>, </title>, <abstract>, </abstract>, <location>, </location>, <abstract>, </abstract>, <biosketch>, </biosketch>, <host>, </host>

Either manually or automatically. XML Schema is weak on semantics.

An Ontology level is needed



- Ontologies add
- Structure
 - Constraints
 - mappings

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

Use Semantic Web Technology to publish shared data & knowledge

Semantic web technologies allow machines to share data and knowledge using common web language and protocols.

~ 1997

Semantic Web beginning

Semantic Web => Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge



2007

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

LOD beginning

Semantic Web => Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge



2008

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

LOD growing

Semantic Web => Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge



2009

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 inter-linked to support integration and fusion of knowledge

2010

...growing faster

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 inter-linked to support integration and fusion of knowledge

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

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 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