

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 knowledge representation, just as the WWW globalize hypertext”

-- Tim Berners-Lee

IOHO

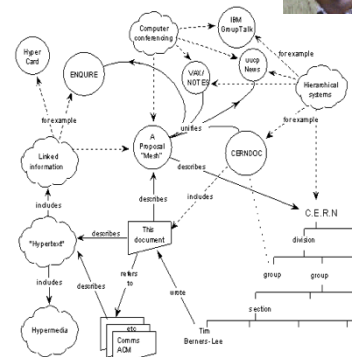
- The web is like a universal acid, eating through and consuming everything it touches.
 - Web principles and technologies are equally good for wireless/pervasive computing
- The semantic web is our first serious attempt to provide semantics for XML sublanguages
- It will provide mechanisms for people 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 (oo)
- W3C' s SW activity (o1)
- W3C' s OWL (o3)



<http://www.w3.org/History/1989/proposal.html>

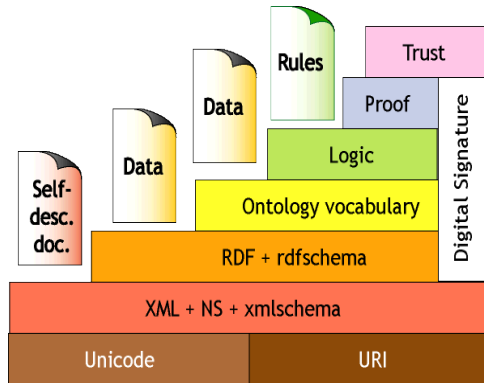
W3C's Semantic Web Goals

Focus on machine consumption:

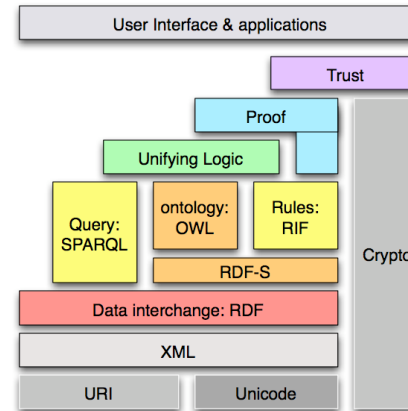
"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

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

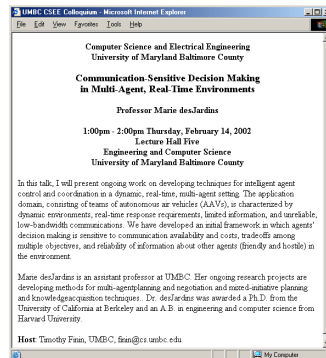
TBL's semantic web vision



Semantic web stack 2006

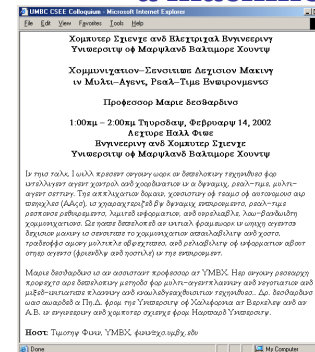


Why is this hard?



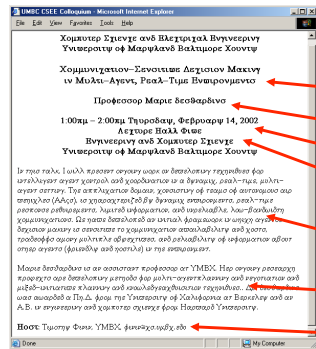
after Frank van Harmelen
and Jim Hendler

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

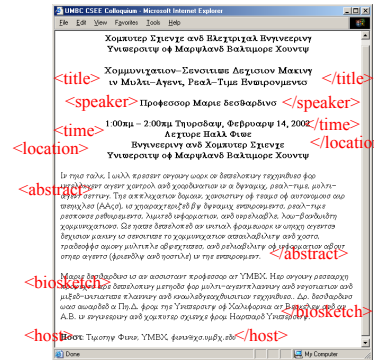
location

abstract

biosketch

host

XML to the rescue?

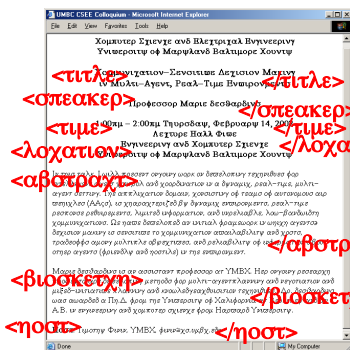


XML fans propose creating a XML tag set to use for each application.

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....

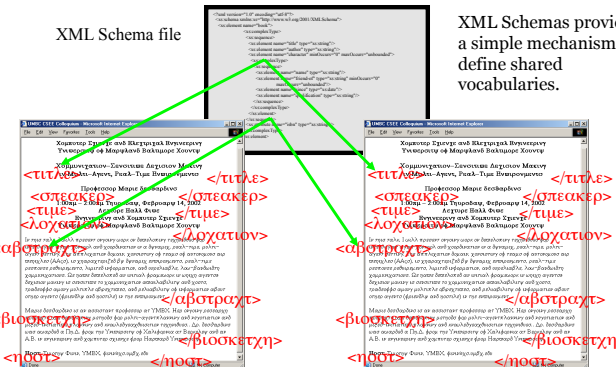
The tag names carry no meaning.

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.

after Frank van Harmelen and Jim Hendler

Semantic Web => Linked Open Data

Use Semantic Web Technology
to publish shared data &
knowledge



2007

Data is inter-
linked to support inte-
gration and fusion of knowledge

LOD beginning

Semantic Web => Linked Open Data

Use Semantic Web Technology
to publish shared data &
knowledge



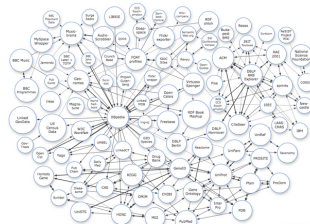
2008

Data is inter-
linked to support inte-
gration and fusion of knowledge

LOD growing

Semantic Web => Linked Open Data

Use Semantic Web Technology
to publish shared data &
knowledge



2009

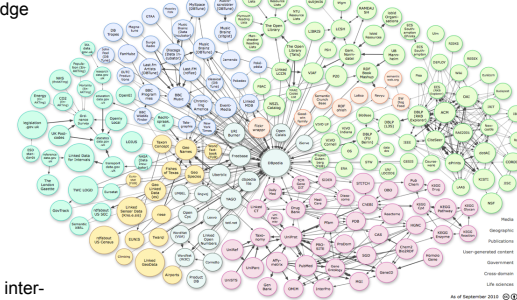
Data is inter-
linked to support inte-
gration and fusion of knowledge

... 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



2010

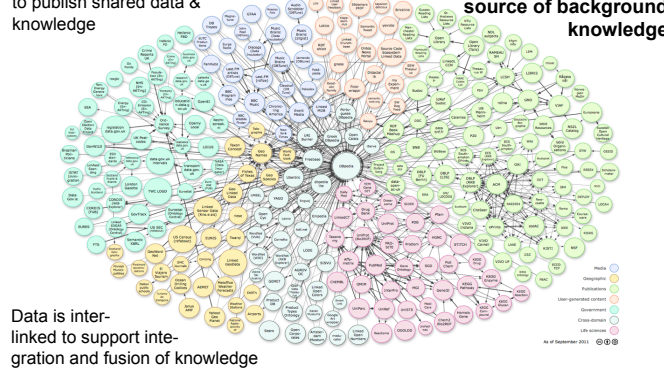
Data is inter-
linked to support inte-
gration and fusion of knowledge

...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 inte-
gration and fusion of knowledge

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

Semantic Web: 1, 2, 3

Traditionally, all languages are divided into three parts:

- 1. Syntax:** legal forms that make up the sentences in a language
- 2. Semantics:** mapping of sentences to meaning (perhaps truth theoretic)
- 3. Pragmatics:** everything else (how to do things with language, knowledge of world, etc.)

1: Syntax

- Use URIs to denote classes, properties, objects, relations
 - http://live.dbpedia.org/resource/Alan_Turing
 - <http://schema.org/Person>
 - <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
- Use strings for literals
- Use triples to make statements
 - `dbpedia:Alan_Turing rdfs:type schema:Person .`
 - “Alan Turing is a Person”

2: Semantics

- Semantics maps URIs to the things they denote in “the world”
- Some of this is in your mind or in how you write your program
- The meaning of some URIs allow automatic inference
 - The parent relation is the inverse of the children relation
 - `schema:parent owl:inverse schema:children`

3: Pragmatics

- Semantics is more than just about truth (statements that assert things)
- We also have to account for commands, requests, questions, context, etc.
 - Some of this is handled by Web protocols (GET, POST)
 - Some by special SQ protocols (e.g., SPARQL for queries and updates)
 - Some by having reference KBs of the world (e.g., Dbpedia) to help identify common entities

Where are we

- The W3C version of the open semantic web has been growing steadily
- The languages and standards are being used in government and industry
 - BBC uses RDF to make up some of its content online
 - Google detects (some) RDF embedded in html pages and exploits it
 - Data.gov has many datasets in RDF

Wikipedia data in RDF

DBpedia

The screenshot shows the DBpedia page for Alan Turing. It lists various properties and their corresponding values in RDF format. For example, 'dbpedia-owl:birthDate' has values '1912-06-23 (xsd:date)' and '1912-06-23 (xsd:date)'. 'dbpedia-owl:birthName' has the value 'Alan Mathison Turing'. 'dbpedia-owl:birthPlace' has values 'dbpedia:Paddington' and 'dbpedia:Maida_Vale'. 'dbpedia-owl:doctoralAdvisor' has values 'dbpedia:King's_College,_Cambridge', 'dbpedia:Princeton_University', 'dbpedia:Royal_Society', 'dbpedia:Order_of_the_British_Empire', 'dbpedia:Fellow_of_the_Royal_Society', and 'dbpedia:Officer_of_the_Order_of_the_British_Empire'. 'dbpedia-owl:doctoralStudent' has values 'dbpedia:Wilmslow', 'dbpedia:Alonzo_Church', 'dbpedia:Robin_Gandy', 'dbpedia:Computer_science', 'dbpedia:Mathematics', and 'dbpedia:Cryptanalysis'. 'dbpedia-owl:field' has values '118802976', 'dbpedia:Turing_machine', 'dbpedia:Cryptanalysis_of_the_Enigma', 'dbpedia:Automatic_Computing_Engine', and 'dbpedia:Turing_test'. 'dbpedia-owl:knownFor' has values 'dbpedia:King's_College,_Cambridge', 'dbpedia:Princeton_University', 'dbpedia:Royal_Society', 'dbpedia:Order_of_the_British_Empire', 'dbpedia:Fellow_of_the_Royal_Society', and 'dbpedia:Officer_of_the_Order_of_the_British_Empire'.

Wikidata

- [Wikidata](#) aims to create a free rdf-like KB about the world that can be read/edited by humans & machines
 - Wikimedia project started in April 2012 with external funding
- Wikidata clients use the repository, e.g., to populate Web pages or Wikipedia infoboxes
- Based on ideas from [Semantic MediaWiki](#) and [Freebase](#)

Open source since 2005

Semantic Media Wiki

Store infobox info in a KB

Semantic MediaWiki (SMW) is a free, open-source extension to MediaWiki – the wiki software that powers Wikipedia – that lets you store and query data within the wiki's pages. Semantic MediaWiki is also a full-fledged framework, in conjunction with many spinoff extensions, that can turn a wiki into a powerful and flexible "collaborative database". All data created within SMW can easily be published via the Semantic Web, allowing other systems to use this data seamlessly.

More about Semantic MediaWiki

- Introduction to SMW
- FAQ
- Talks and publications
- Tutorials

Installation

- Semantic MediaWiki 1.8
- Administrator manual
- Installation
- Configuration
- Related extensions

User community

- Getting support
- SMW Community Wiki
- List of SMW-using wikis
- SMWCon, the Semantic MediaWiki Conference

The SMW project

- About the project
- Version history
- Development roadmap
- Programmer's guide

News

- 10 January 2013: Survey: What do you think of SMW?

Acquired by Google in 2010

Freebase

Alan Turing, OBE, FRS (/ˈtʊrɪŋ/ TEWR-ing; 23 June 1912 – 7 June 1954) was a British mathematician, logician, cryptanalyst, and computer scientist. He was highly influential in the development of computer science, giving a formalisation of the concepts of "algorithm" and "computation" with the Turing machine, which can be considered a model of a general purpose computer. Turing is widely considered to be the father of computer science. More

Related Topics

- Edward Thomas Hall
- Alan Turing
- William Kingdon Clifford
- Ada Lovelace

People

Place of birth: Maida Vale, United Kingdom

Alan Turing Quotes

cc *Mathematical reasoning may be regarded...*

cc *No, I'm not interested in developing a powerful brain...*

"An entity graph of people, places and things, built by a community that loves open data"

Google Knowledge Graph

The Knowledge Graph

Learn more about one of the key breakthroughs behind the future of search.

See it in action

Discover answers to questions you didn't know to ask and explore collections of related information.

Google's slogan for the knowledge graph: "things, not strings"

Map "mention strings" to entities

Knowledge Graph

Uses data from Freebase

alan turing

Web Images Maps Shopping News More Search tools

30 personal results, 4,130,000 other results.

Alan Turing - Home Page
www.turing.org.uk/turing/
Alan Turing Home Page. Guide to a large website maintained by Andrew Hodges, biographer of Alan Turing (1912-1954).

Alan Turing - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Alan_Turing
Alan Mathison Turing, OBE, FRS was a British mathematician, logician, cryptanalyst, and computer scientist. He was highly influential in the development of ...
Turing machine - Turing test - Halting problem - Logic

BBC - History - Alan Turing (pictures, video, facts & news)
www.bbc.co.uk/history/people/alan_turing
Alan Turing was an English mathematician, wartime code-breaker and pioneer of computer science. Photo: Alan Turing with two colleagues and a Faraday.

BBC News - Alan Turing: 'house's suicide verdict' not supportable
www.bbc.co.uk/news/science-environment-15681052
Jun 28, 2012 - Alan Turing, the British mathematical genius and codebreaker, may not have committed suicide, as is widely believed, claims an academic.
George Willman shared this on Google+

Alan M. Turing (British mathematician and logician) - Britannica
www.britannica.com/EBchecked/topic/609730/Alan-M-Turing
British mathematician and logician, who made major contributions to mathematics, cryptanalysis, logic, philosophy, and biology and to the new areas later ...

The Alan Turing Year - 2012 Turing Centenary

Alan Turing
Alan Mathison Turing, OBE, FRS, was a British mathematician, logician, cryptanalyst, and computer scientist.
VRs: 100
Born: June 23, 1912, Maida Vale
Died: June 7, 1954, Wilmslow
Books: Computing machinery and intelligence, More
Education: Princeton University (1936 - 1938), More
Parents: Ethel Sara Stoney, Julius Mathison Turing
Siblings: John Turing

Google's slogan for the knowledge graph: "things, not strings"

Annotate your web pages in RDFa => object in the FB graph

Facebook Open Graph

The Open Graph protocol

Introduction

The Open Graph protocol enables any web page to become a rich object in a social graph. For instance, this is used on Facebook to allow any web page to have the same functionality as any other object on Facebook.

While many different technologies and schemas exist and could be combined together, there isn't a single technology which provides enough information to richly represent any web page within the social graph. The Open Graph protocol builds on these existing technologies and gives developers one thing to implement. Developer simplicity is a key goal of the Open Graph protocol which has informed many of the technical design decisions.

Basic Metadata

To turn your web pages into graph objects, you need to add basic metadata to your page. We've based the initial version of the protocol on RDFa which means that you'll place additional <meta> tags in the <head> of your web page. The four required properties for every page are:

- og:title - The title of your object as it should appear within the graph, e.g., "The Rock".
- og:type - The type of your object, e.g., "video.movie". Depending on the type you specify, other properties may also

speech => text => entities => task SIRI engineers from AI/SW community

Apple's SIRI

Siri.
Your wish is its command.

Siri lets you use your voice to send messages, schedule meetings, place phone calls, and more.* Ask Siri to do things just by talking the way you talk. Siri is so easy to use and does so much, you'll keep finding more and more ways to use it.

SIRI needs lots of semantic data about entities in the world

IBM's Watson

IBM Watson: Ushering in a new era of computing

What's next for Watson?
We want to hear how YOU would put Watson to work. Join the discussion.

Putting Watson to work
Data is growing at extraordinary rates - 800

Take the next step to your business ready for Watson? Discover how to prepare your

Recent news

- IBM hopes its Watson will become doctor's
- IBM Exploring Ways to Use, Sell Watson Cloud Services
- More Watson news

IBM used Semantic Web technology and data in Watson, see <http://bit.ly/X44aIE>

Summary

- The Web's made people smarter by letting us share information and knowledge as text, audio and images
- Machines should also be able to use the web to publish and retrieve information and knowledge
- Human forms of knowledge are hard for machines to understand and generate
- The Semantic Web is a collection of languages, ontologies, software tools, services and KBs that are designed to support machines