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

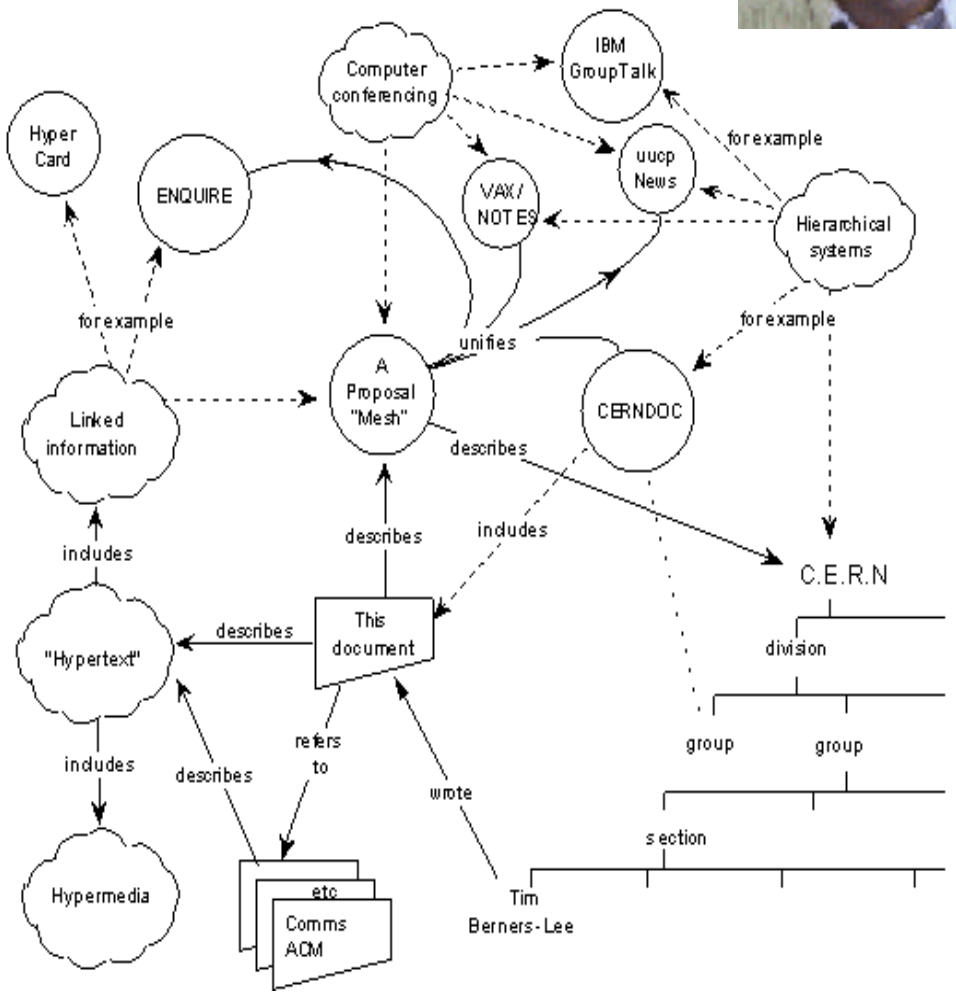
- The web is like a universal acid, eating through and consuming everything it touches.
  - Web principles and technologies are equally good for wireless/ pervasive computing
- The semantic web is our first serious attempt to provide semantics for XML sublanguages
- It will provide mechanisms for people 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 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

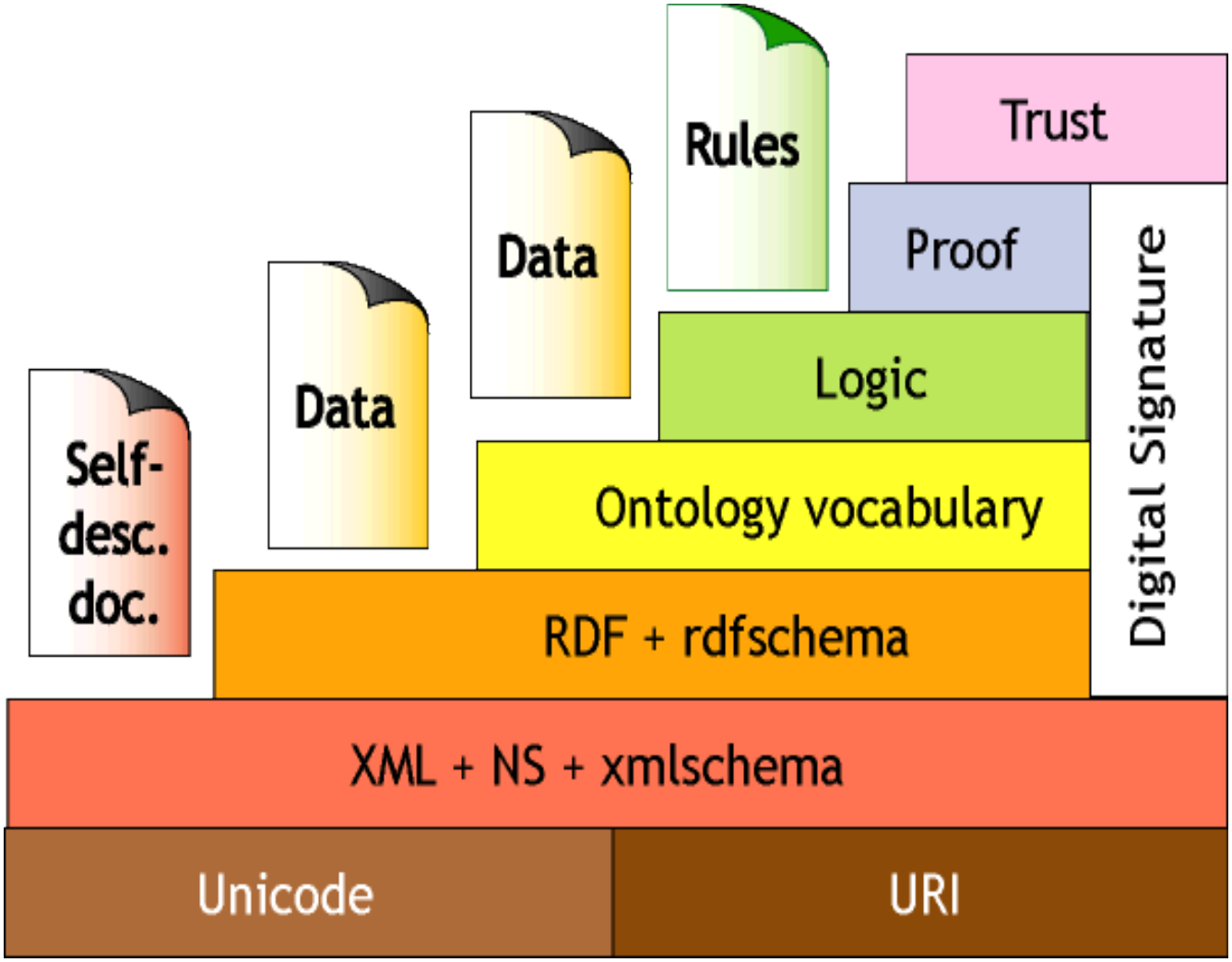
W3C's Semantic Web Goals

Focus on machine consumption:

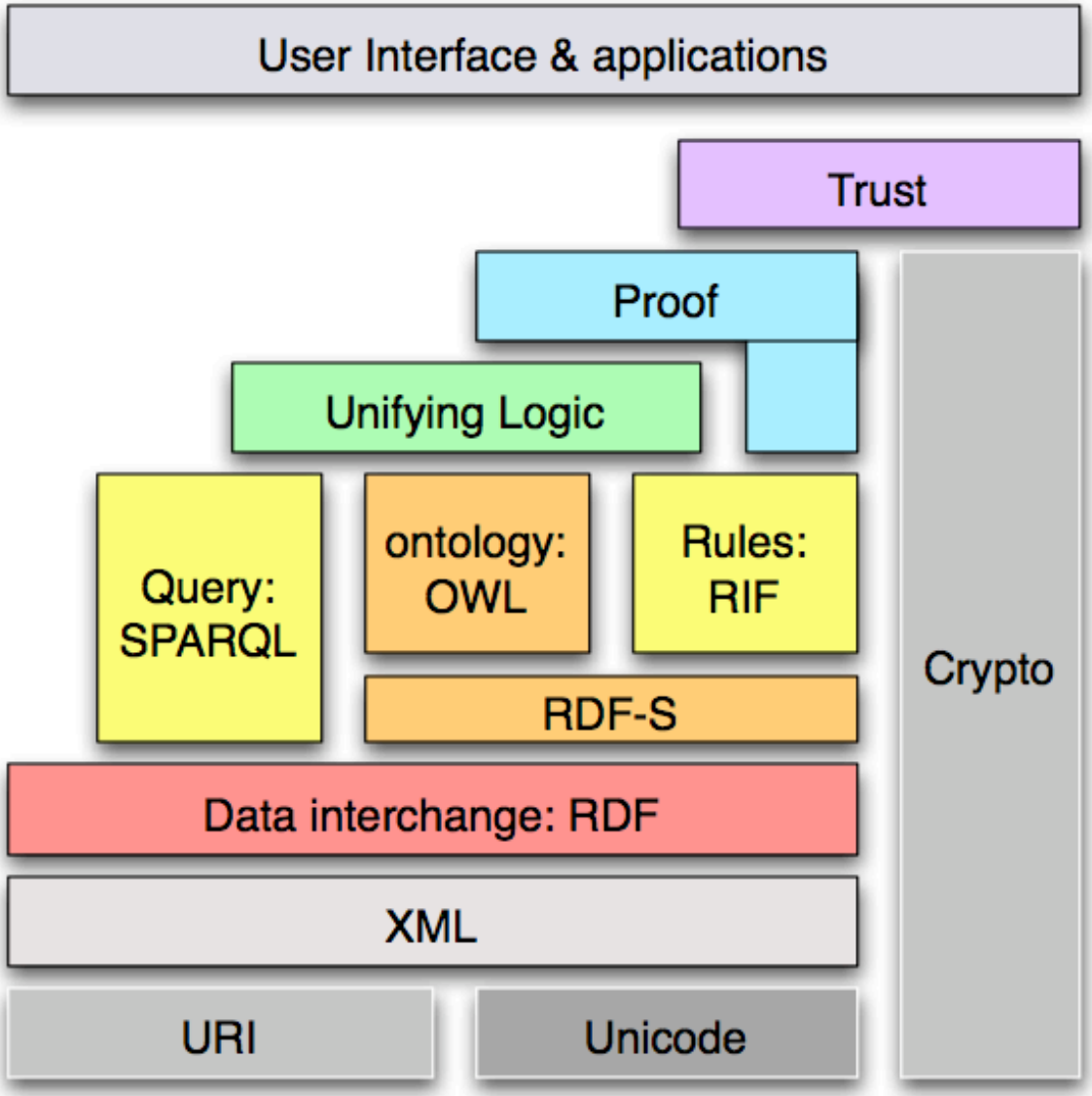
"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

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

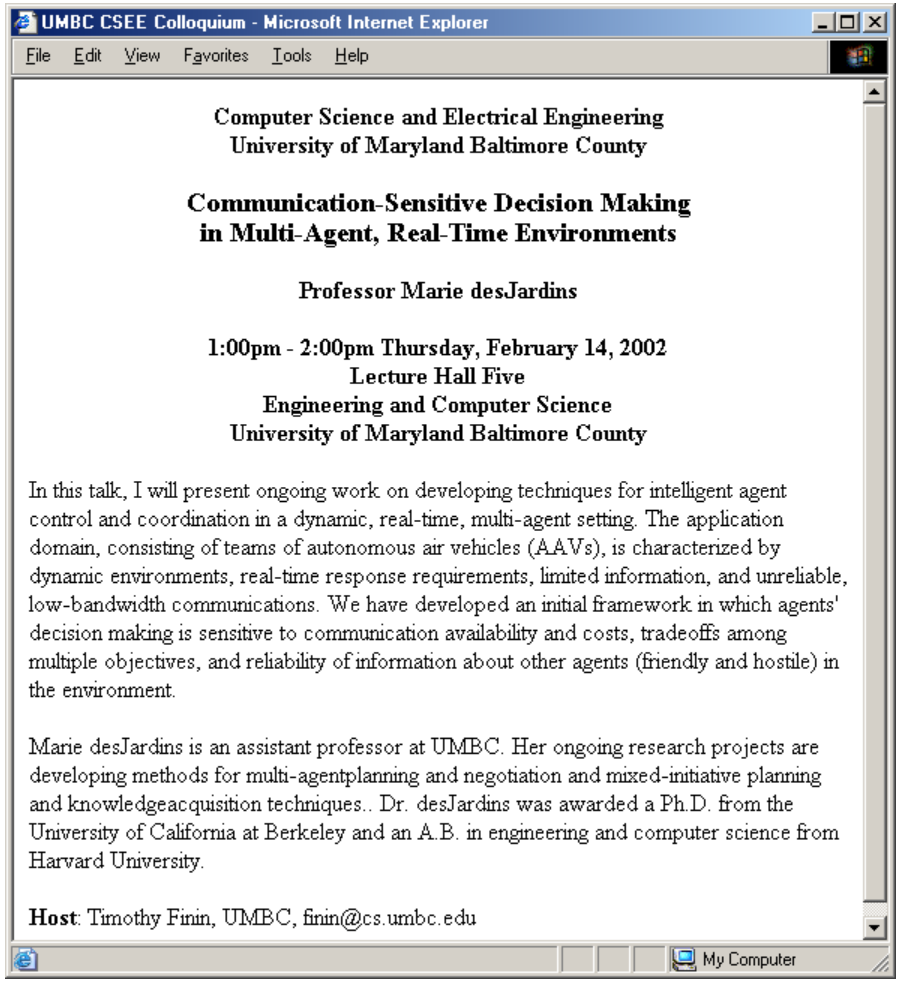
TBL's semantic web vision



Semantic web stack 2006

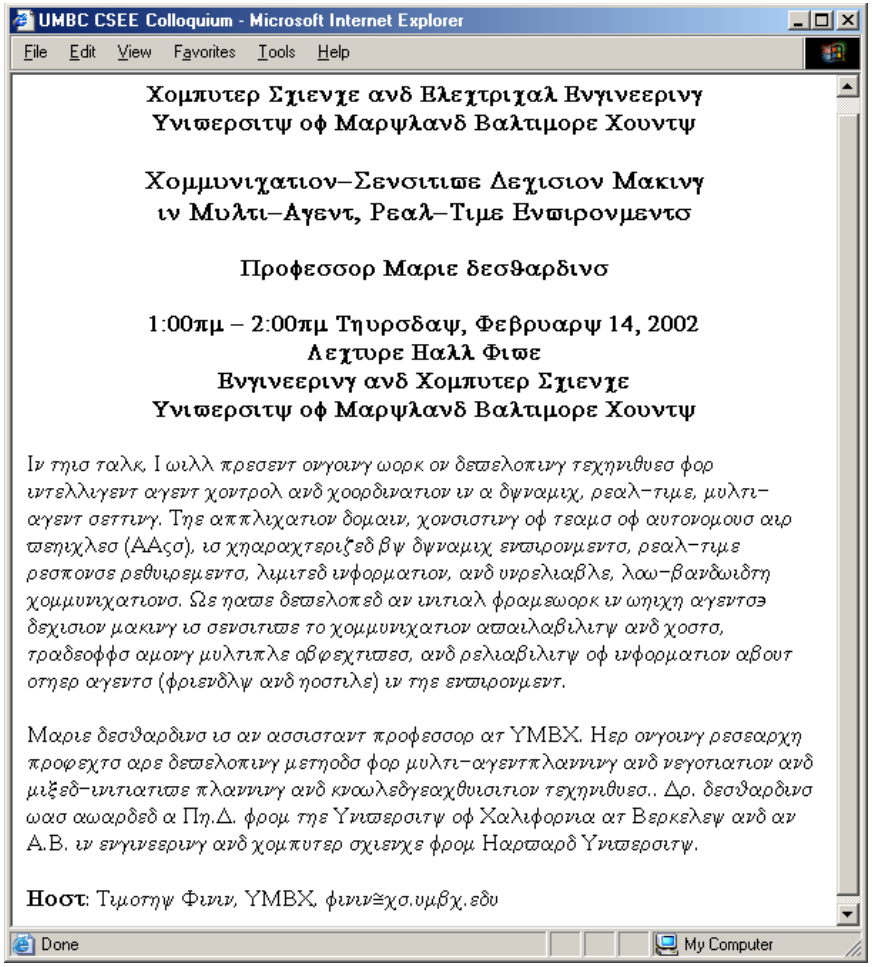


Why is this hard?



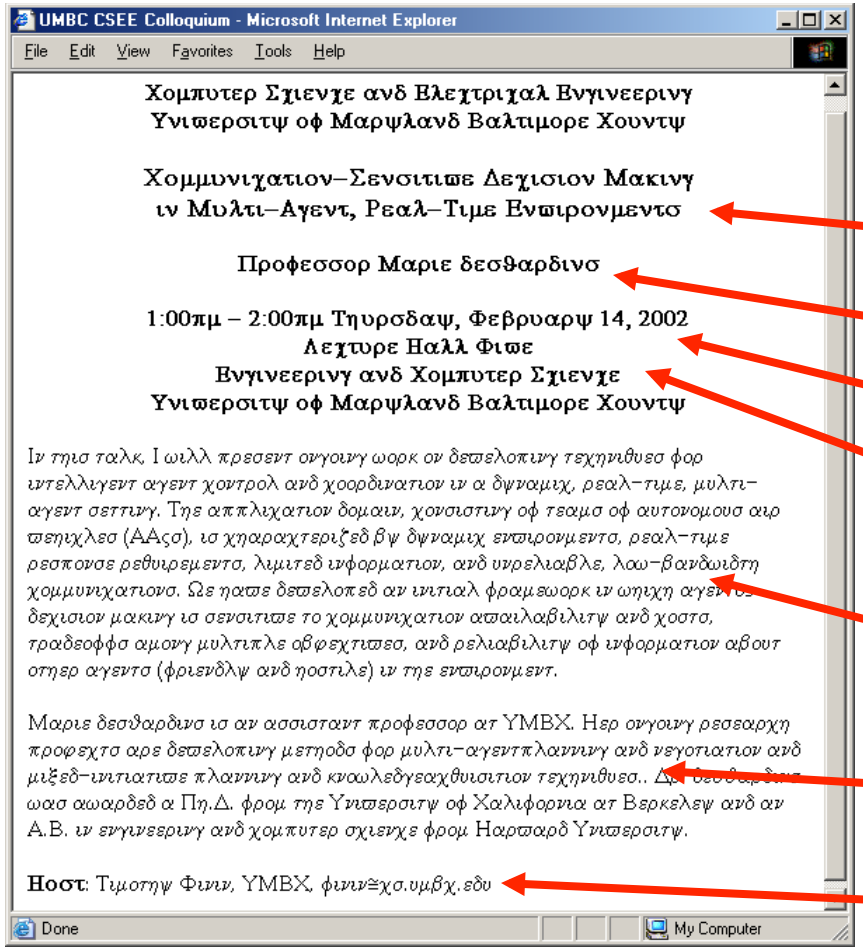
after Frank van Harmelen and Jim Hendler

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

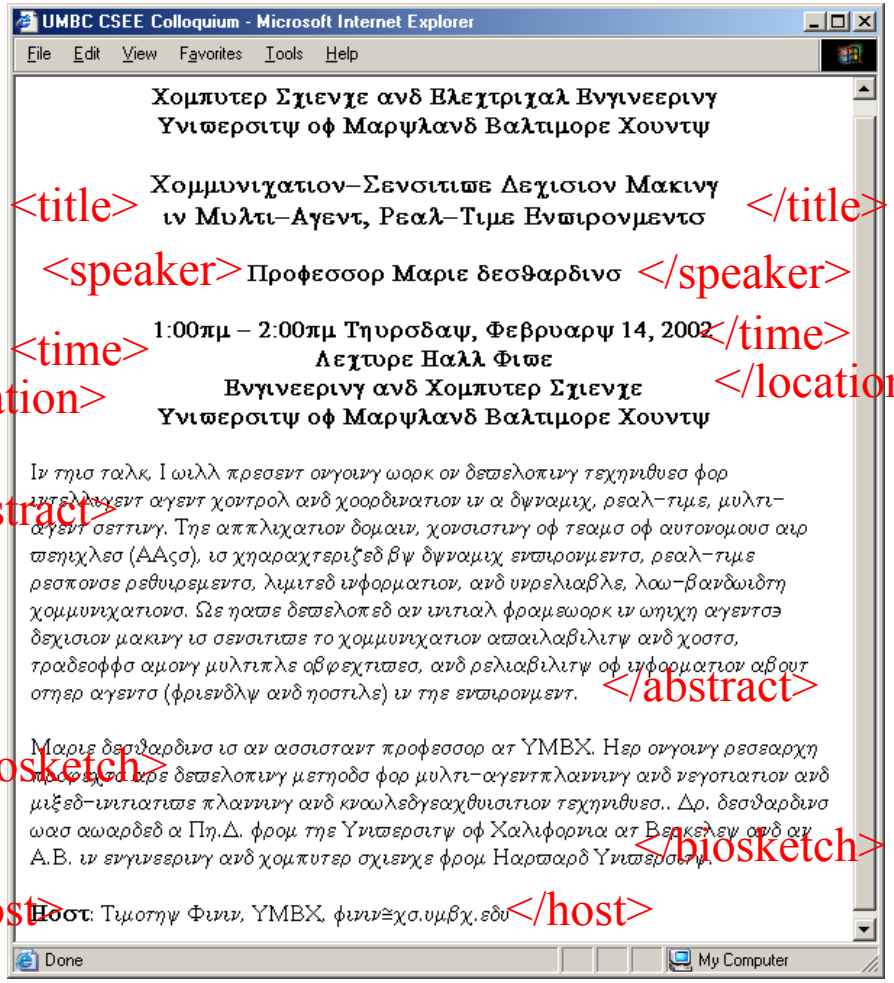
location

abstract

biosketch

host

XML to the rescue?

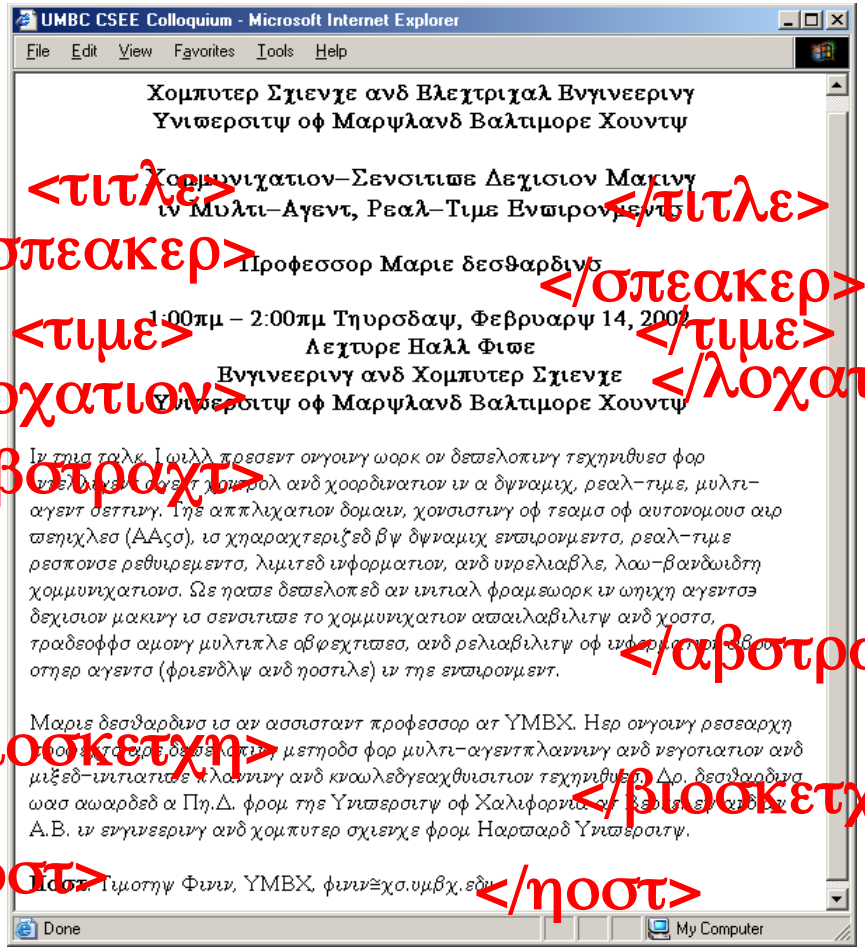


XML fans propose creating a XML tag set to use for each application.

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….

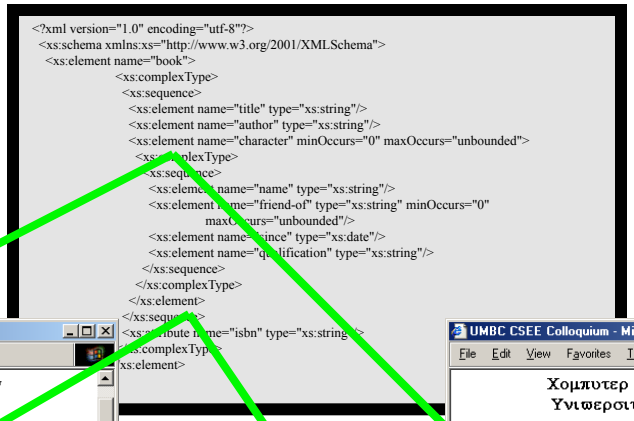
The tag names carry no meaning.

XML DTDs and Schemas have little or no semantics.

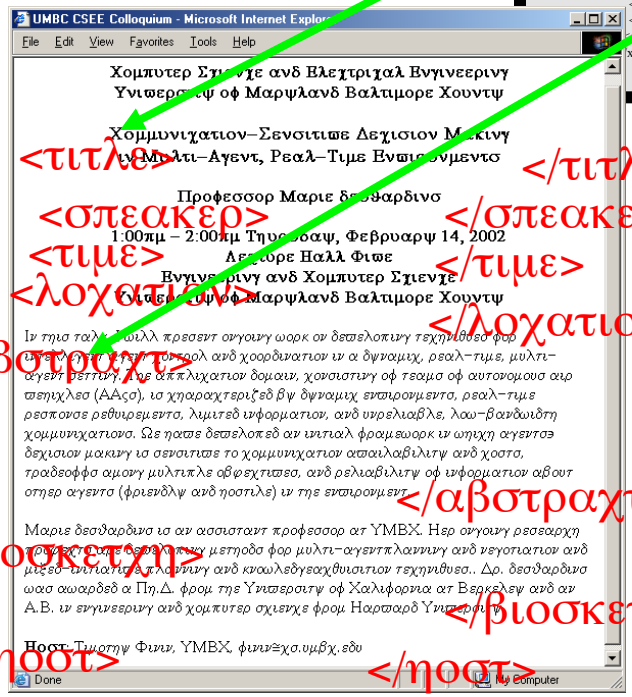
after Frank van Harmelen and Jim Hendler

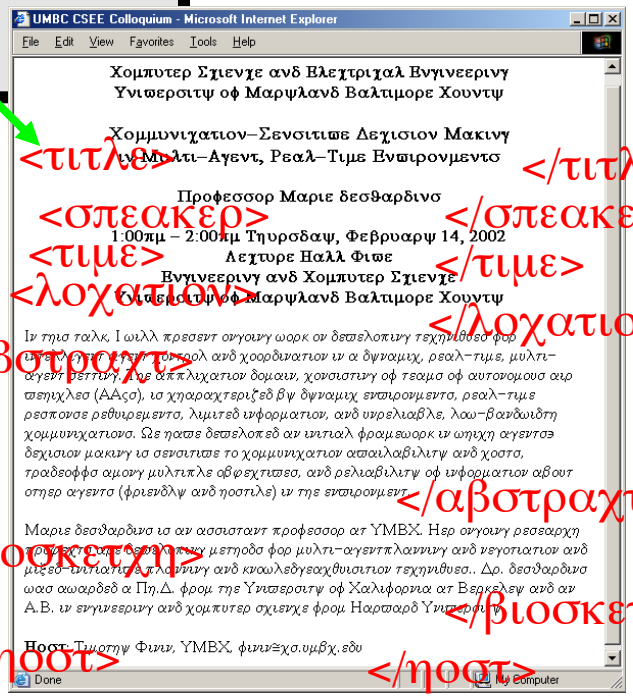
XML Schema helps

XML Schema file



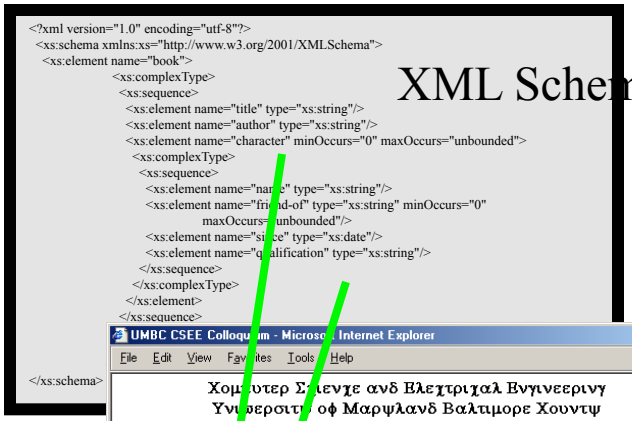
XML Schemas provide a simple mechanism to define shared vocabularies.

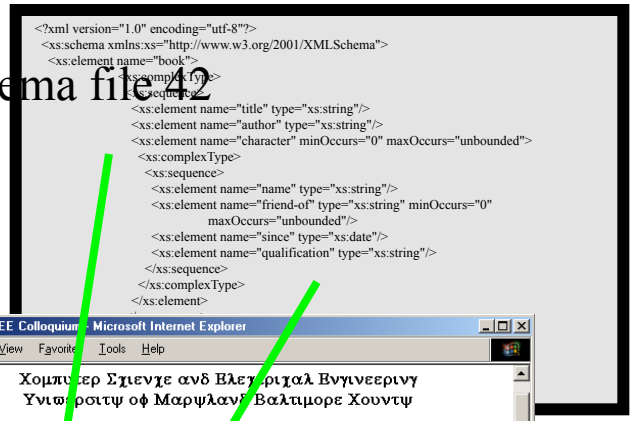


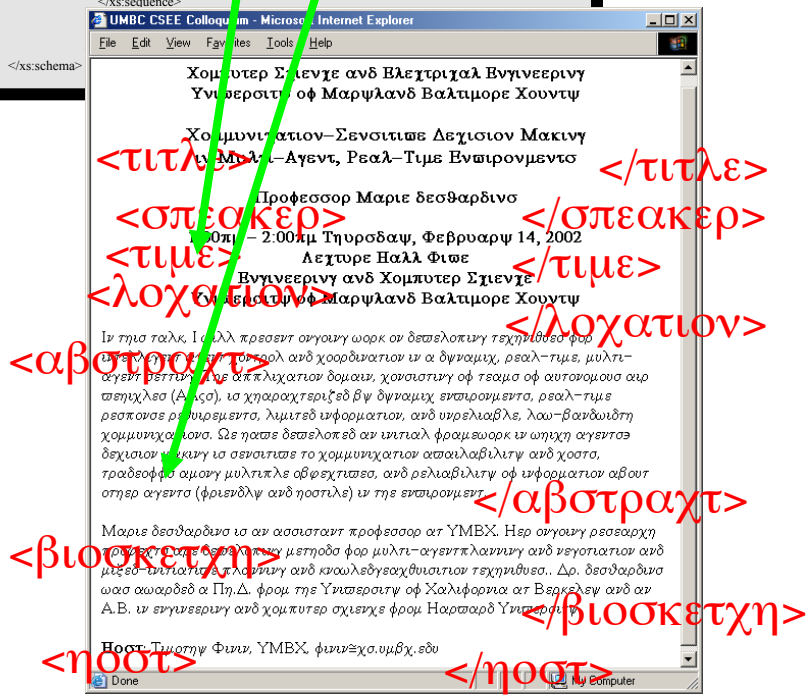


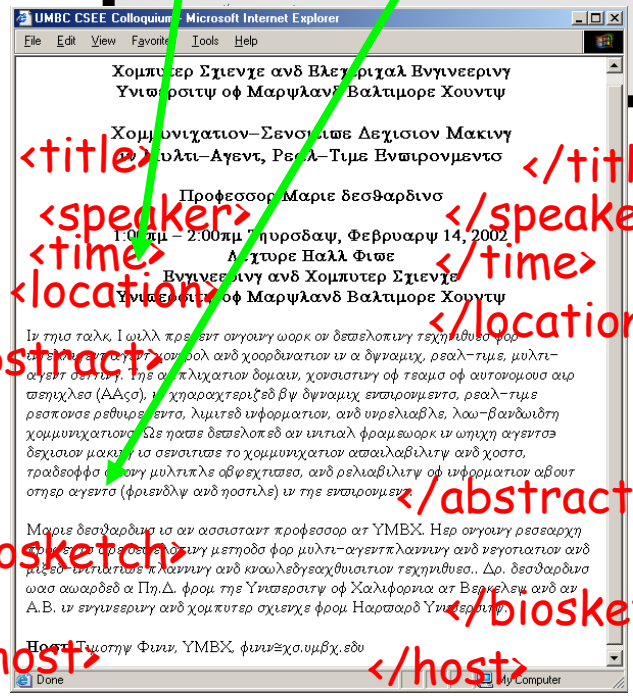
after Frank van Harmelen and Jim Hendler

But there are many schemas



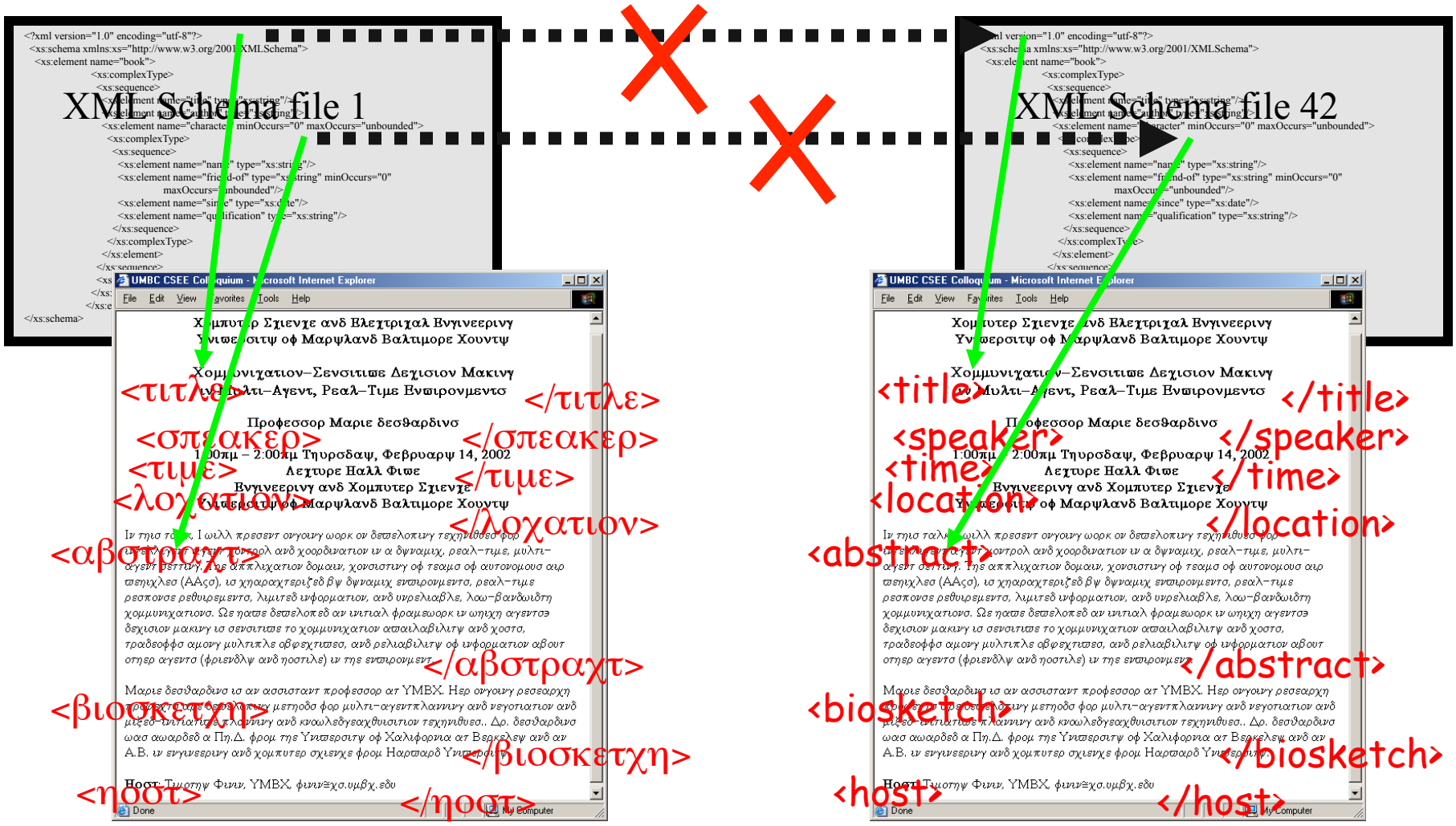






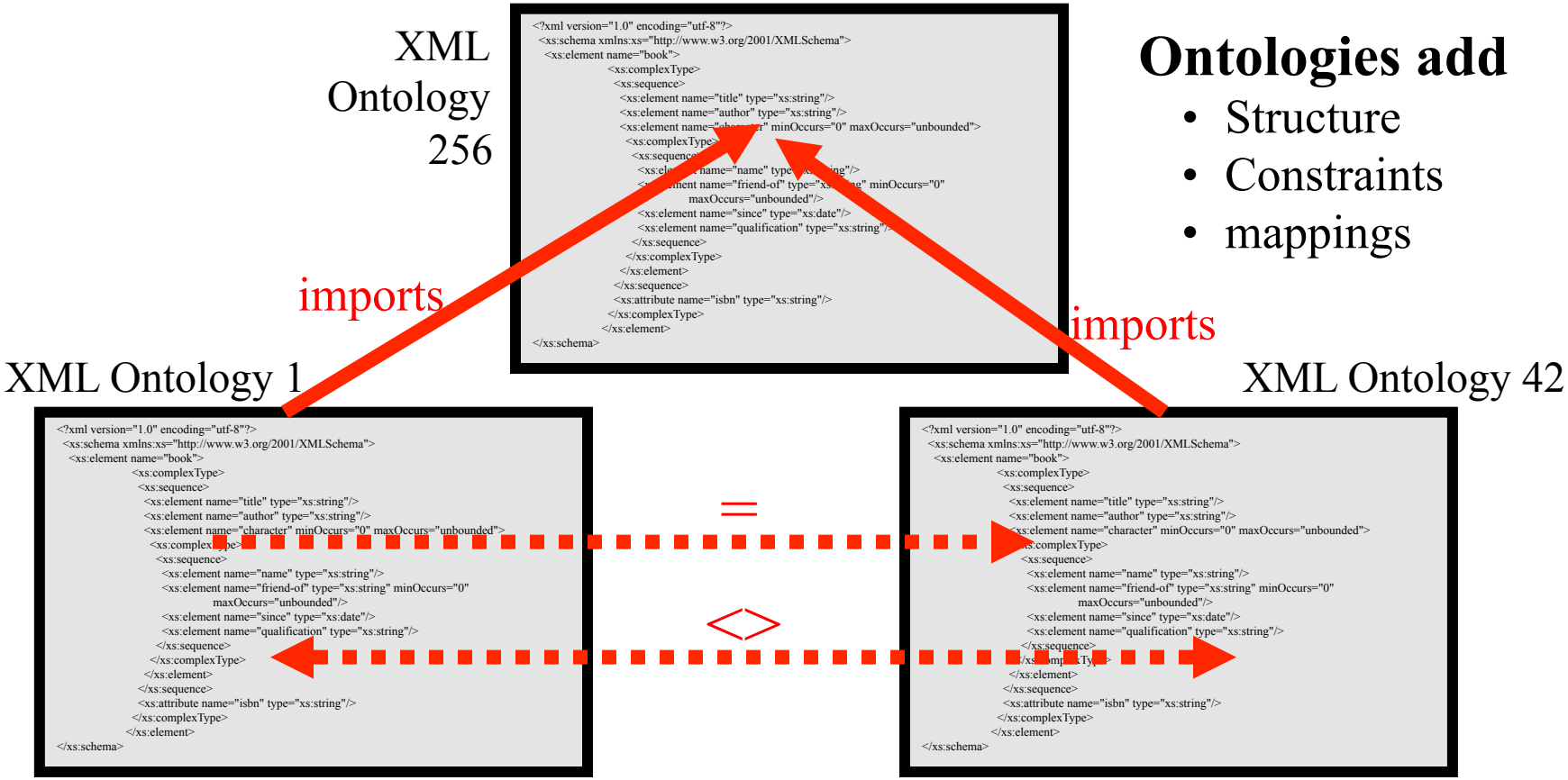
after Frank van Harmelen and Jim Hendler

There's no way to relate schema



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

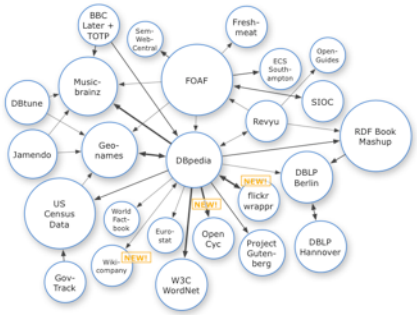
Semantic web technologies
allow machines to share
data and knowledge using
common web language and
protocols.

~ 1997

Semantic Web beginning

Semantic Web => Linked Open Data

Use Semantic Web Technology to publish shared data & knowledge



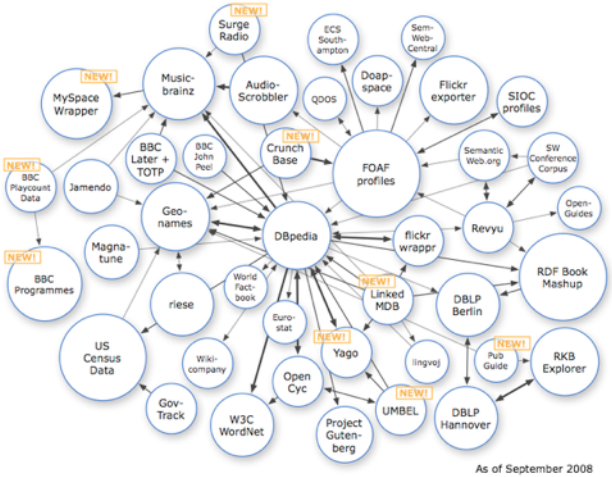
2007

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

LOD beginning

Semantic Web => Linked Open Data

Use Semantic Web Technology
to publish shared data &
knowledge



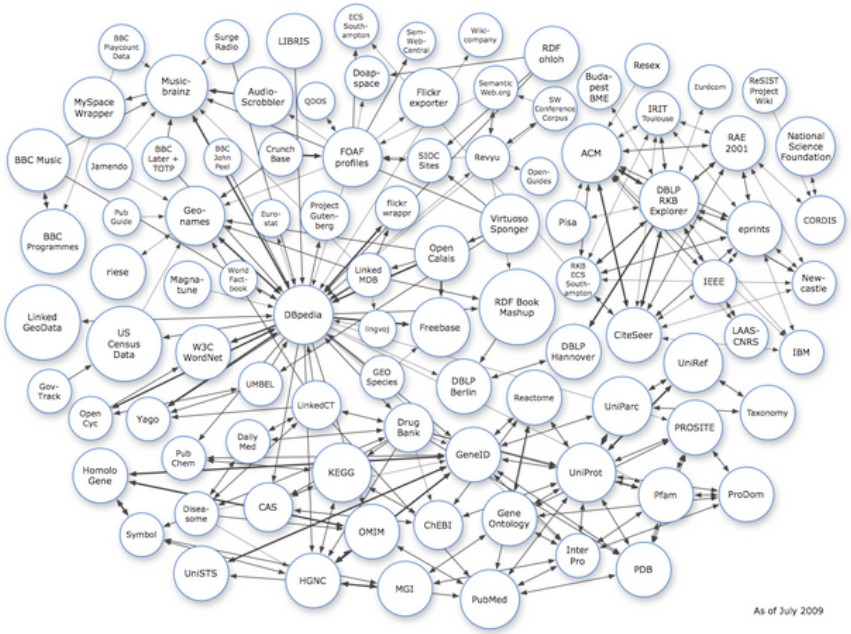
2008

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

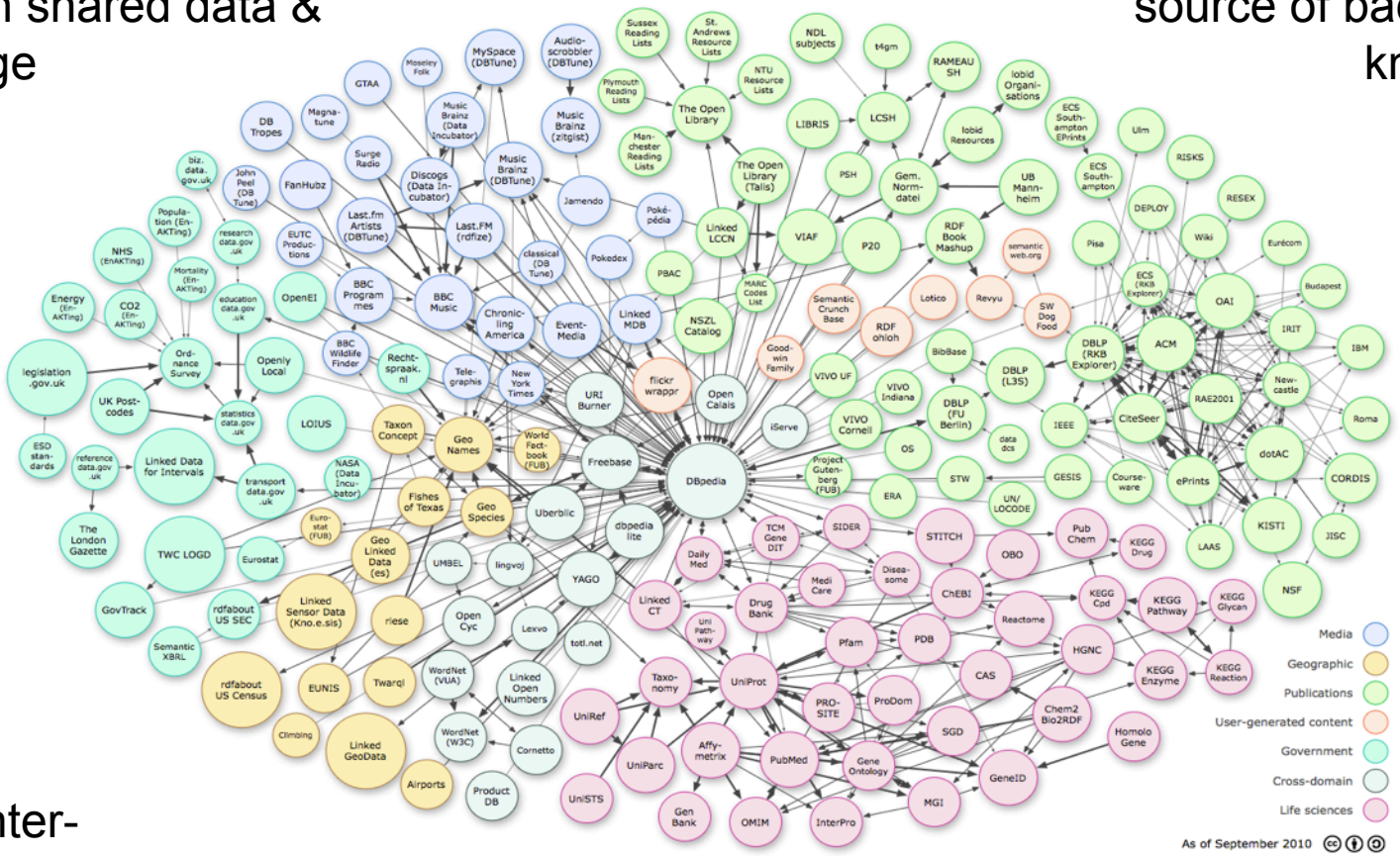
Data is inter-
linked to support inte-
gration and fusion of knowledge

… 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

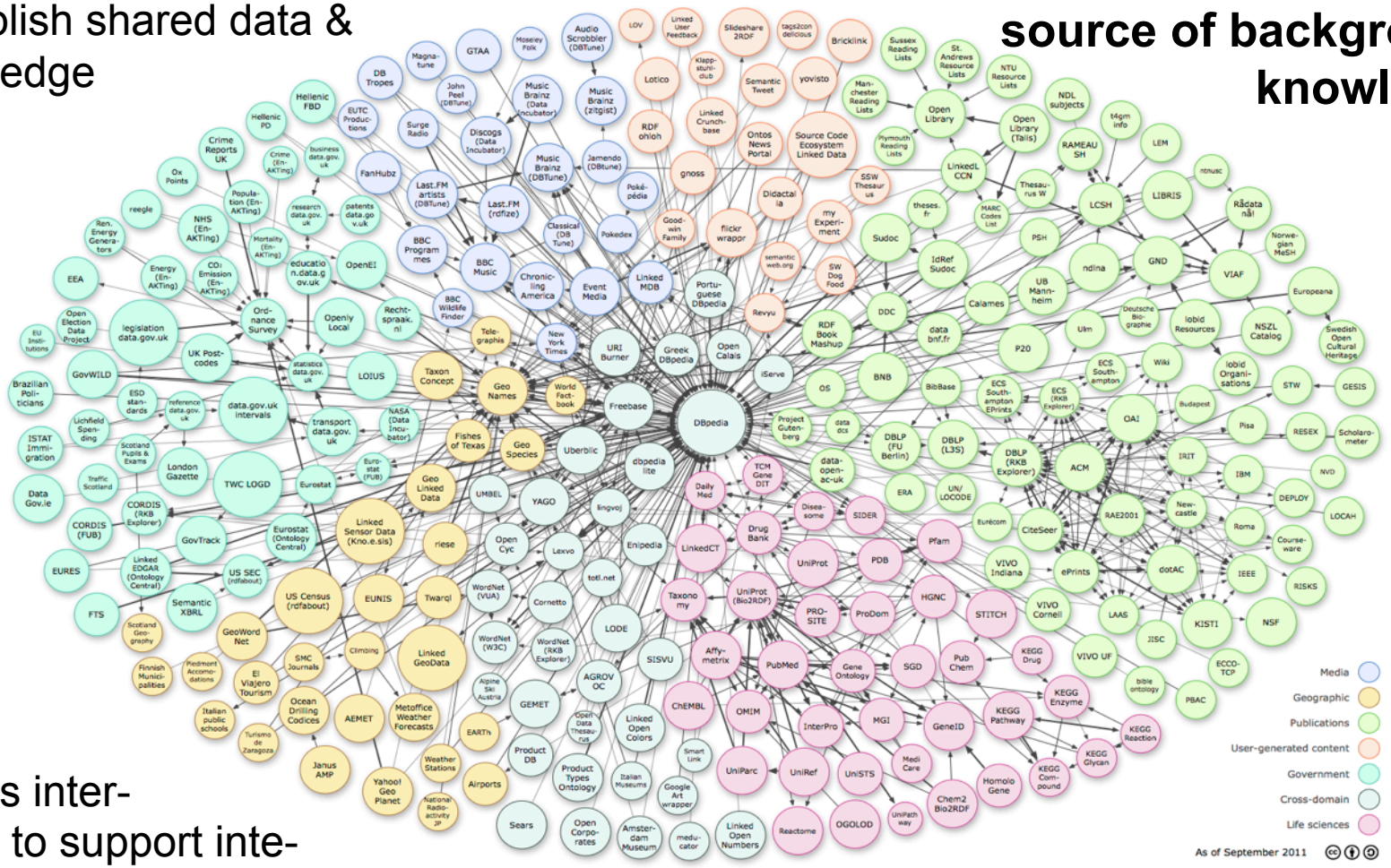
Data is inter-linked to support inte-gration and fusion of knowledge

…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 in 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

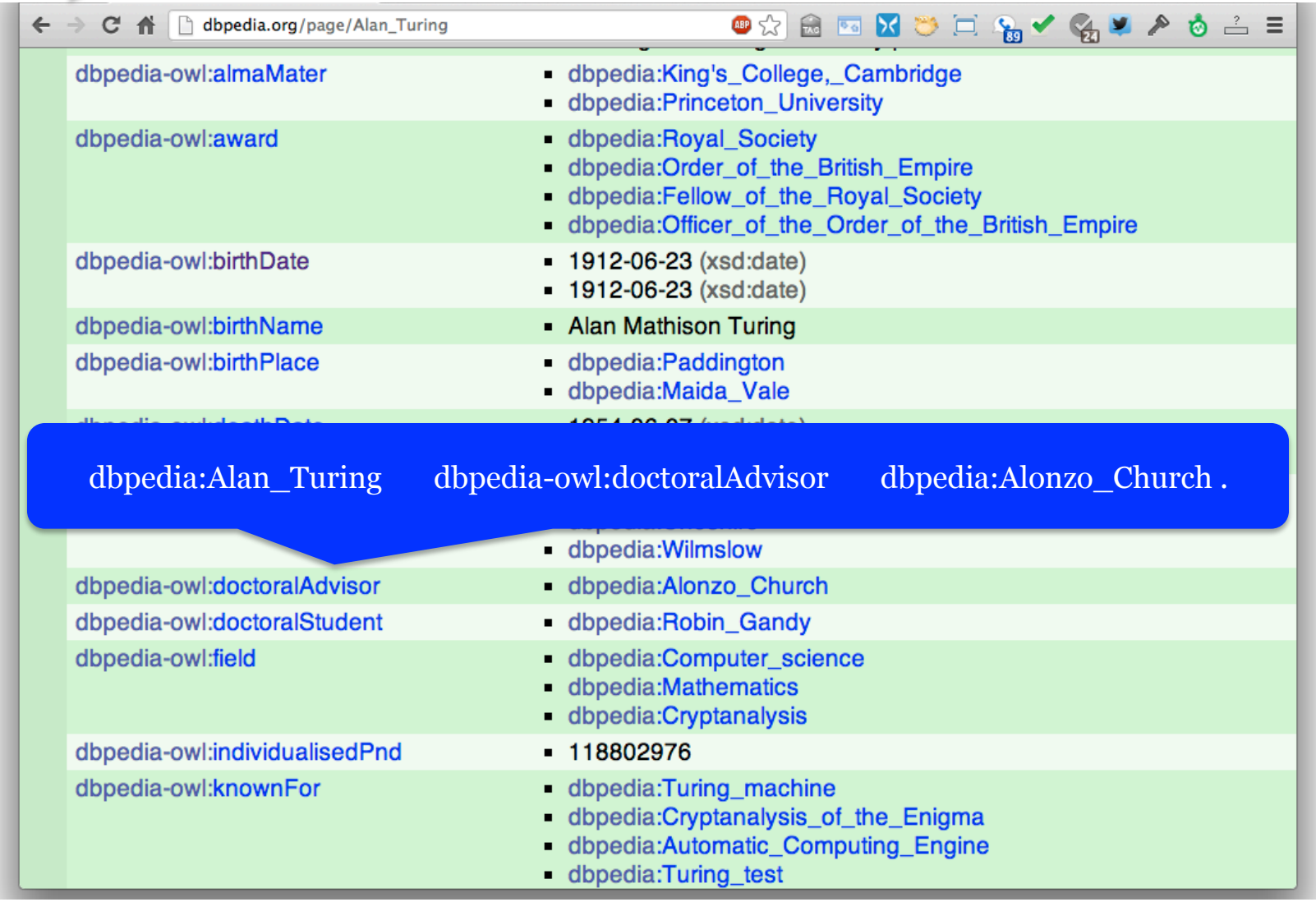
- Semantics is more than just about truth (statements that assert things)
- We also have to account for commands, requests, questions, context, etc.
  - Some of this is handled by Web protocols (GET, POST)
  - Some by special SQ protocols (e.g., SPARLQ for queries and updates)
  - Some by having reference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



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

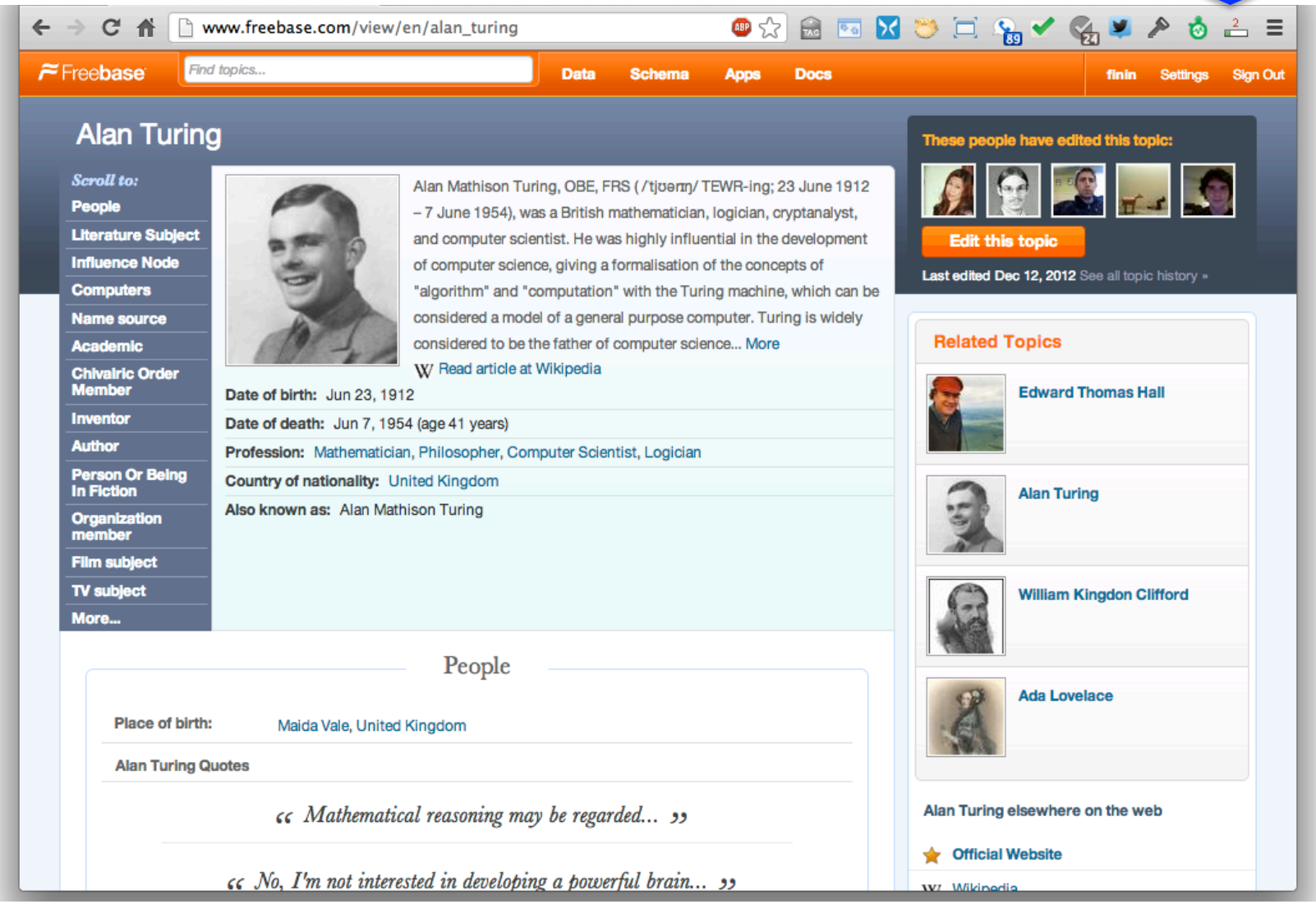
Open source since 2005

Semantic Media Wiki

Store infobox info in a KB

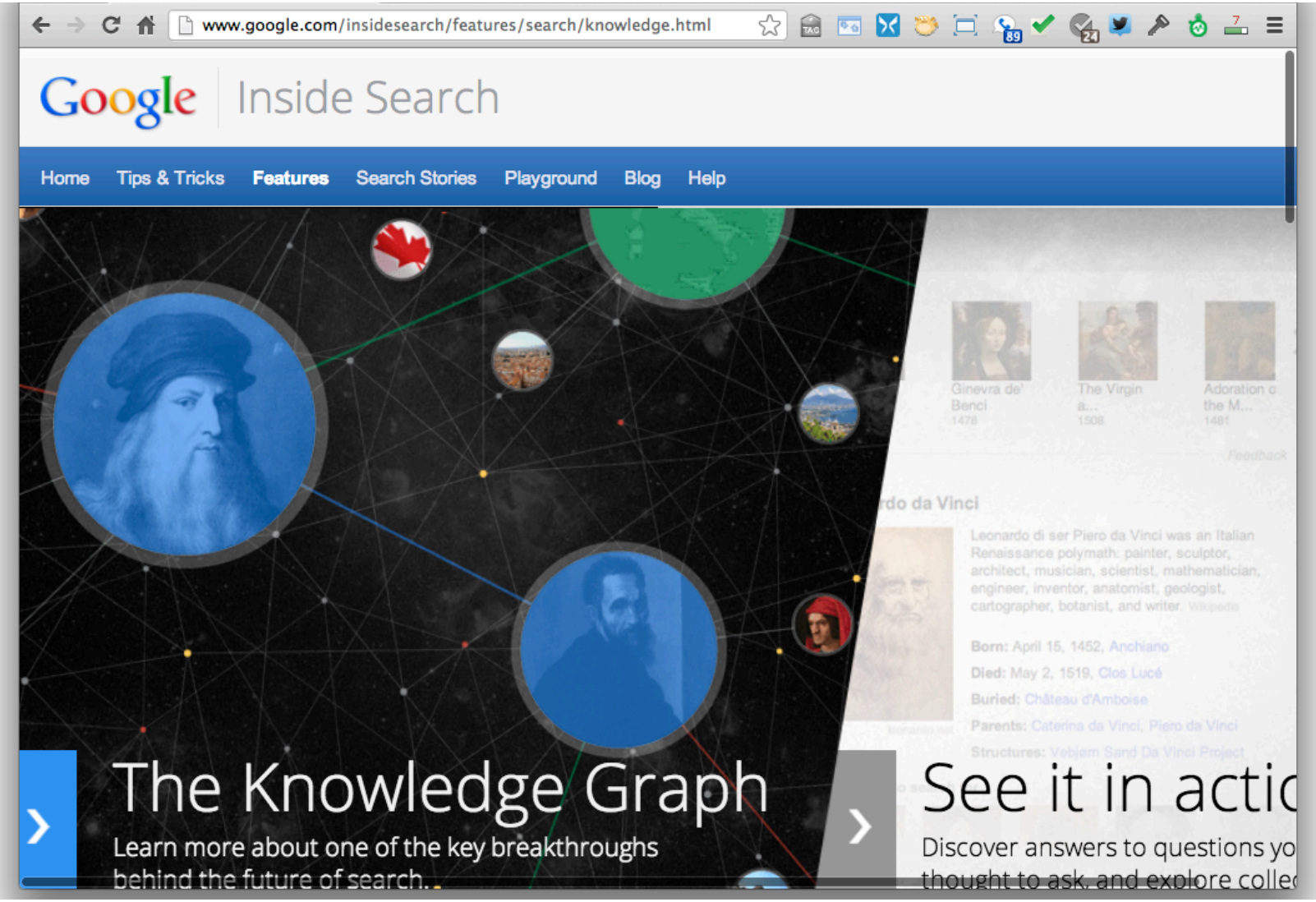
Freebase

Acquired by Google in 2010



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

Google Knowledge Graph

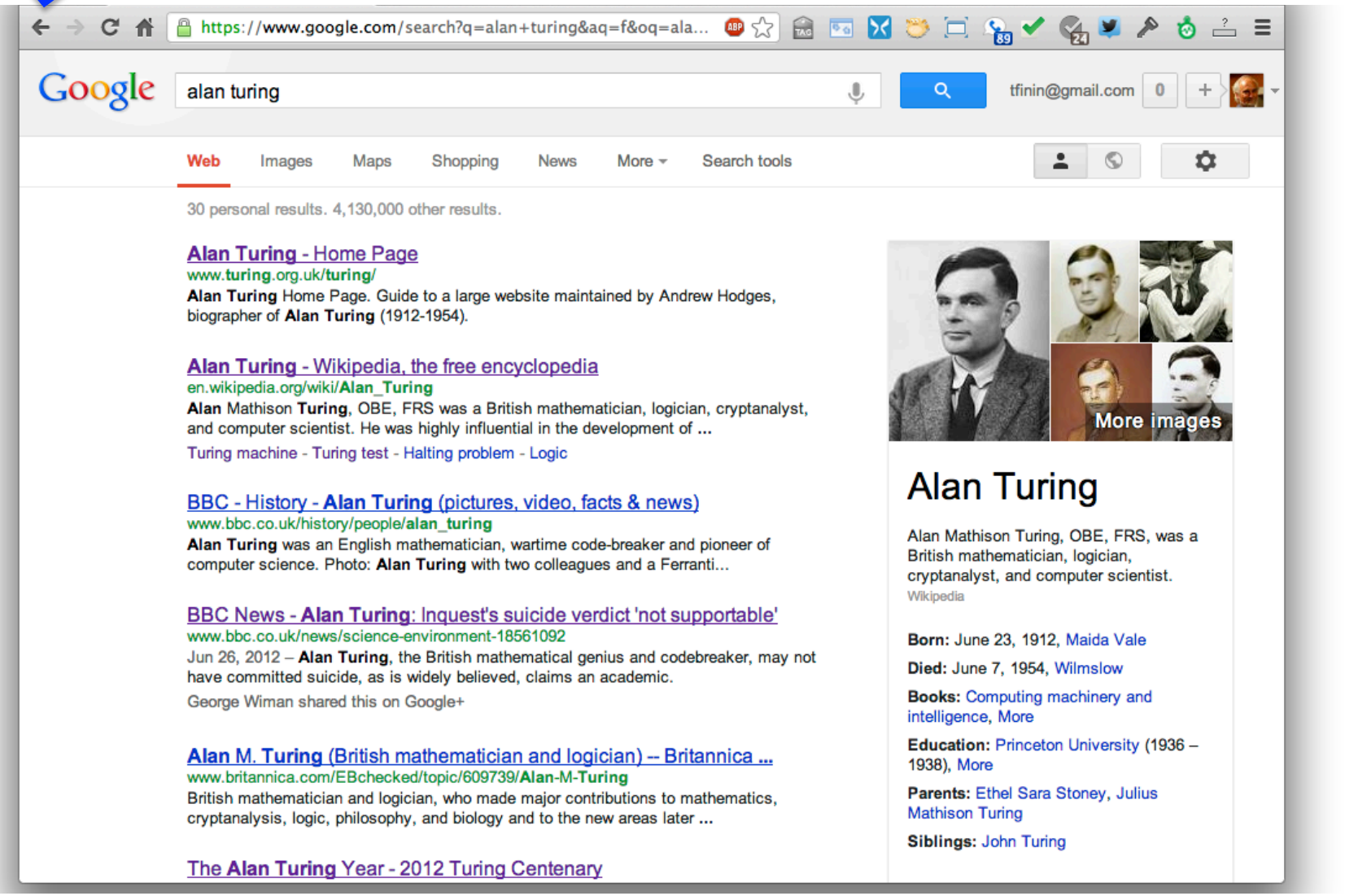


Google's slogan for the knowledge graph: "things, not strings"

Map "mention strings" to entities

Knowledge Graph

Uses data from Freebase

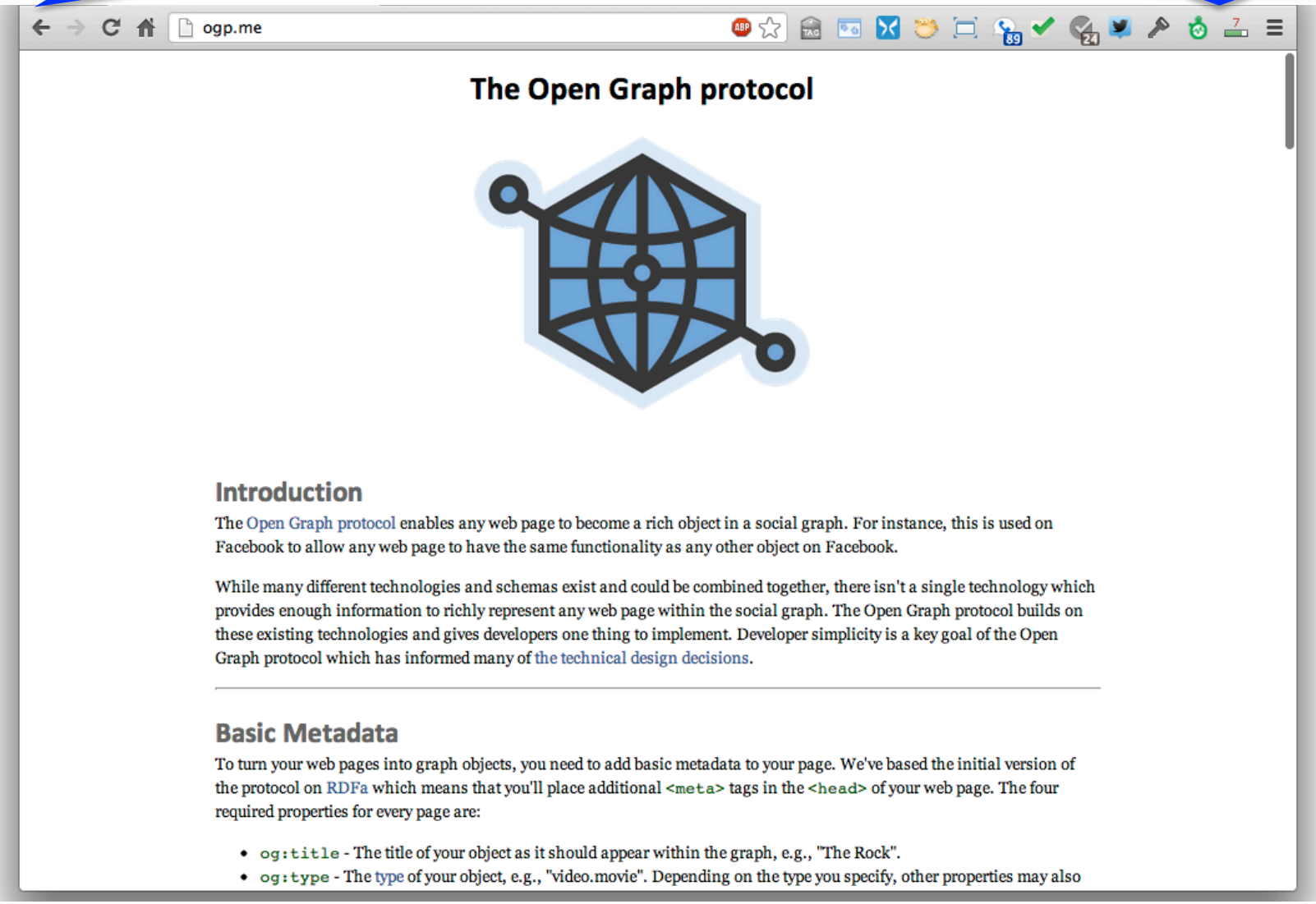


Google's slogan for the knowledge graph: "things, not strings"

Annotate your web pages in RDFa

Facebook Open Graph

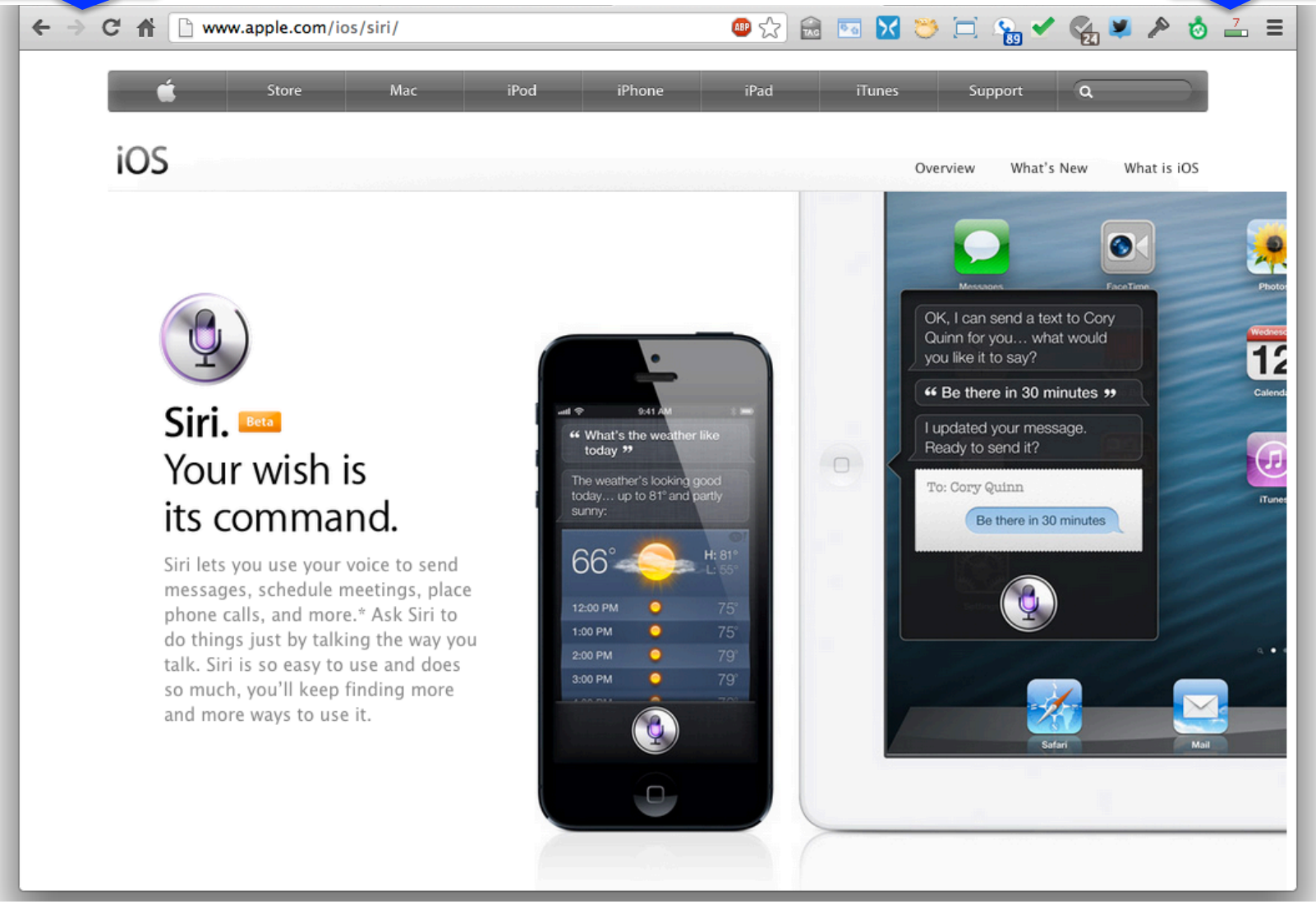
=> object in the FB graph



speech => text => entities => task

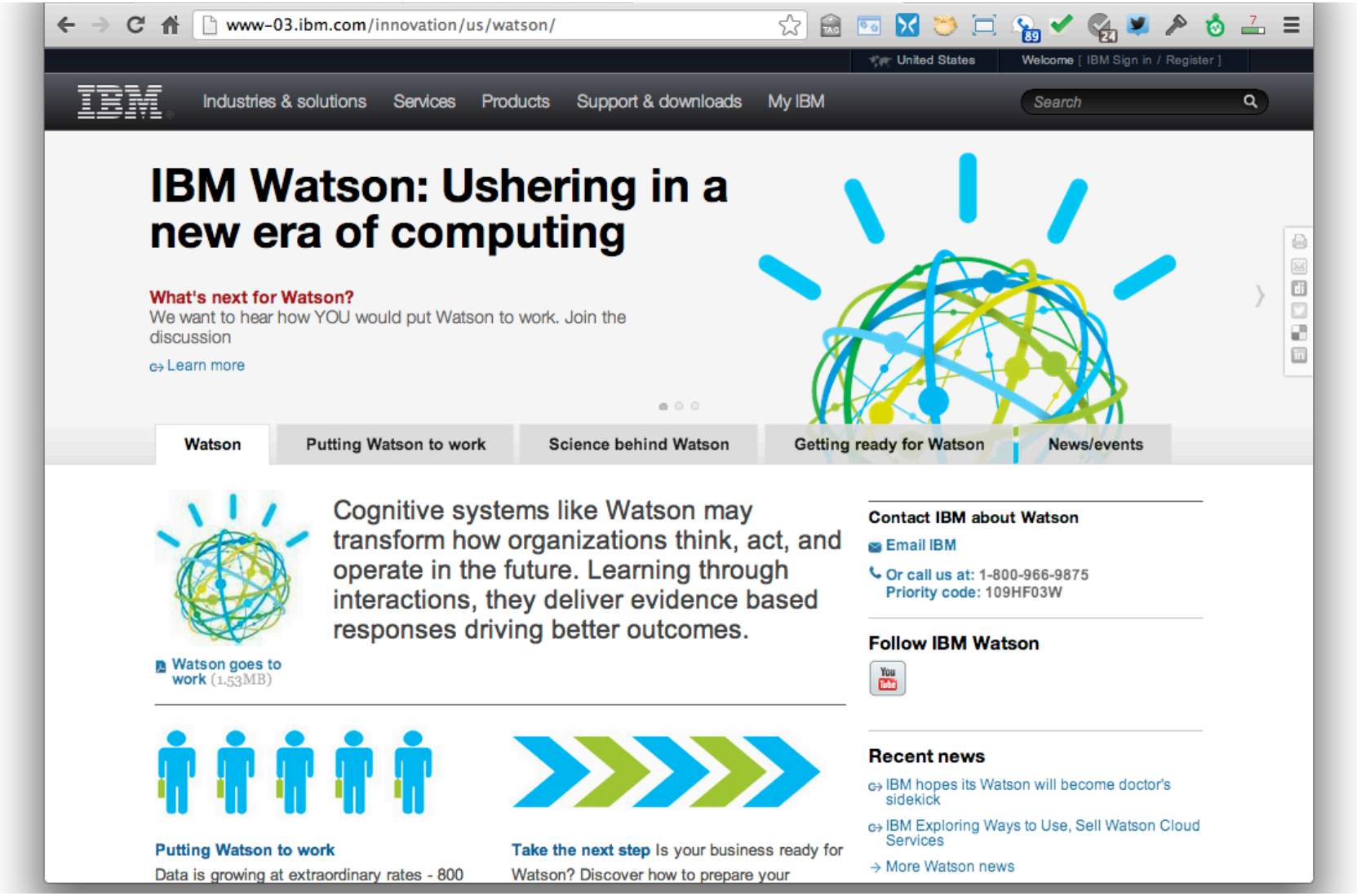
Apple's SIRI

SIRI engineers from AI/SW community



SIRI needs lots of semantic data about entities in the world

IBM's Watson



IBM used Semantic Web technology and data in Watson, see http://bit.ly/X44alE

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