Web-based Knowledge Representation

Stefan Schlobach, Eyal Oren

Vrije Universiteit Amsterdam

4th February 2009
How big is the Web?
How many sites?
How many people?
How old is the Web?
How old is the Internet?
The Web is huge. Lots of data. Can you keep it consistent?

No.
The Web is huge. Lots of data. Can you keep it consistent?

No.

Can you use the information?

Yes.
The Web is huge. Lots of data. Can you keep it consistent?

No.

Can you use the information?

Yes.

Can your machine use the information?

The Web is huge. Lots of data. Can you keep it consistent?

No.

Can you use the information?

Yes.

Can your machine use the information?


Let’s have a look . . .
Vannevar Bush (1945): Memex
Doug Engelbart (1960): NLS
Ted Nelson (1965): Hypertext
Tim Berners-Lee (1990): http://info.cern.ch
Netscape IPO (1994): $4.4 billion
Google IPO (2004): $1.67 billion
From Computer Networks to the Web of Things

- Usenet email: allegro!batcave!cornell!rpics!weltyc

Internet: “it’s not the wires, but the computers”

Web: “it’s not the computers, but the documents”

Semantic Web: “it’s not the documents, but the things”

We’re interested in the knowledge on the Web, not the documents
From Computer Networks to the Web of Things

▶ Usenet email: allegro!batcave!cornell!rpics!weltyc
▶ Internet: “it’s not the wires, but the computers”
From Computer Networks to the Web of Things

- Usenet email: allegro!batcave!cornell!rpics!weltyc
- Internet: “it’s not the wires, but the computers”
- Web: “it’s not the computers, but the documents”
From Computer Networks to the Web of Things

- Usenet email: allegro!batcave!cornell!rpics!weltyc
- Internet: “it’s not the wires, but the computers”
- Web: “it’s not the computers, but the documents”
- Semantic Web: “it’s not the documents, but the things”
From Computer Networks to the Web of Things

- Usenet email: allegro!batcave!cornell!rpics!weltyc
- Internet: “it’s not the wires, but the computers”
- Web: “it’s not the computers, but the documents”
- Semantic Web: “it’s not the documents, but the things”

We’re interested in the knowledge on the Web, not the documents
The Web for Humans

A city, with population, flag, geo-location
The Web for Machines

Characters and tables, but what does it mean?
Why can’t we query all the information on the Web as a database?
Data Interoperability on the Web

- Different formats, different structures, different vocabularies, different concepts, different meaning
- Data should be structured (not ASCII)
- Structure should be data-oriented (not HTML)
- Meaning of data should be clear (not XML)
- Data should have standard APIs (not Flickr)
- Reusable mappings between data are needed (not XSLT)
“There is lots of data we all use every day, and it’s not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

No. Why not? Because we don’t have a web of data. Because data is controlled by applications and each application keeps it to itself.”

Tim Berners-Lee, W3C Semantic Web
The Web of Data

- Musicbrainz
- Wiki-company
- Gov-Track
- US Census Data
- World Factbook
- FOAF
- DBpedia
- Freshmeat
- BBC Later + TOTP
- Sem-Web-Central
- ECS Southampton
- SW Conference Corpus
- SIOC
- Onto-world
- Freshmeat
- Open-Guides
- New!
- Open-Cyc
- Project Gutenberg
- W3C WordNet
- Statista
- EU Open Data
- DBLP Berlin
- DBLP Hannover
- DBLP
- Revyu
- RDF Book Mashup
- Later + TOTP
- Freshmeat
- New!
- New!
- New!
- New!
- New!
What can you do with the Web of Data?
Which Ajax players were born in Amsterdam?
Which Dutch soccer player plays in a stadium >40,000 people?
What is the Semantic Web (2)?

- **RDF**: basic data format (triples form hypergraph)
  
  ```
  john likes mary .
  ```

- **RDFS**: simple schema language (subclass, subproperty)
  
  ```
  dc:title rdfs:subPropertyOf rdfs:label .
  Jeep isa Car .
  ```

- **OWL**: rich schema language (constraints, relations)
  
  ```
  likes isa owl:symmetricProperty .
  ```

- RDF allows interlinked graphs of arbitrary structure
- RDFS and OWL allow inferencing of implicit information
Overview of the WebKR course

- How to represent, query, and reason with Web knowledge?
- How to express basic facts?
  Feb 11th: RDF(S)
- How to query Web data?
  Feb 13th: SPARQL
- How to express richer ontologies?
  Feb 18th: OWL
- How to manipulate and convert (non-Semantic) Web data?
  Feb 20: XML
Does the Semantic Web really exist?
The state of the Semantic Web

- Stable formats and standards: RDF, RDFS, OWL, SPARQL
- Technology: Yahoo, Adobe, Oracle, IBM, HP, Software AG
- Deployment: Novartis, Pfizer, Telefonica, Vodafone, Elsevier
- Data: Gene ontology, Geonames, Uniprot, Wordnet, DBLP
- Applications visibility: Twine, Vodafone Live, Freebase, Garlik

[Ivan Herman, W3C]
Practicalities

- Two lectures per week until March, then final assignment.
- Lecture assignment after each lecture: pass/fail.
- Hand-in before next lecture (see assignment).
- Pass 4 out of 5: bonus point.
- Do not hand-in by email unless the assignment says so.

- Final assignment: groups of two.

- Book: “a Semantic Web primer”.
- Blackboard and dedicated mailing list.
- No class on Friday 6th Feb.