Introduction to
Semantic Web

Erdogan Dogdu

TOBB University of Economics and Technology

Outline

- Motivation
  - Development of the Web: Web 1.0, 2.0, 3.0
  - Limitations of the current Web
- Technical solution
  - Introduction to Semantic Web
  - Semantic Web – architecture and languages
  - Semantic Web - data
  - Semantic Web – tools, processes
- Recent trends
- Summary
- References

Semantic web

- What?
  - Web vs Internet
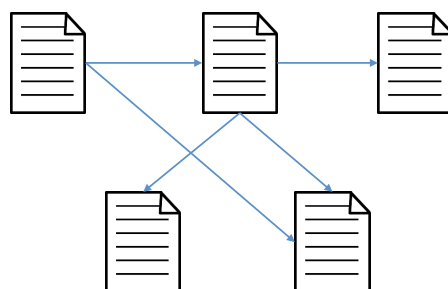 - Semantics?
- How?

Before the Web

Memex

- "As We May Think", Vannevar Bush, 1945, *The Atlantic Monthly*
- **Mem(**ory-Ind)**ex**
- All information (books, communication, etc) in the form of microfilms and indexed using electromechanical device.

Web 1.0 (1990)

- Hyperlinked web documents
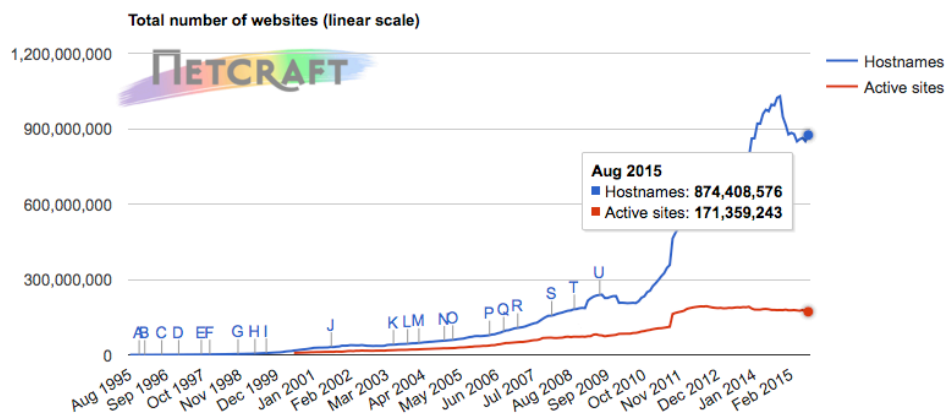

Abstracts away…
- Storage
- Networking

Web 2.0 (2000)

- Applications and content
- Social web
  - Facebook, Twitter, Google+, LinkedIn
  - YouTube, Flickr
  - Blogs, wikis
  - Wikipedia
- Problem
  - Application silos
  - Disconnected data
  - Big data

Web is big!

#Web pages in billions...

WorldWideWebSize.com
DAILY ESTIMATED SIZE OF THE WORLD WIDE WEB

The size of the World Wide Web (The Internet)
The Indexed Web contains **at least 3.32 billion pages** (Thursday, 10 July, 2014).
The Dutch Indexed Web contains **at least 244.19 million pages** (Thursday, 10 July, 2014).

The size of the World Wide Web (The Internet)

The Indexed Web contains **at least 4.79 billion pages** (Monday, 07 September, 2015).
The Dutch Indexed Web contains **at least 359 million pages** (Monday, 07 September, 2015).

BİL 546 - Semantic Web – TOBB ETÜ

9

How many docs w/«the»?

Google   the

Web   Books   News   Images   Maps   More   Search tools
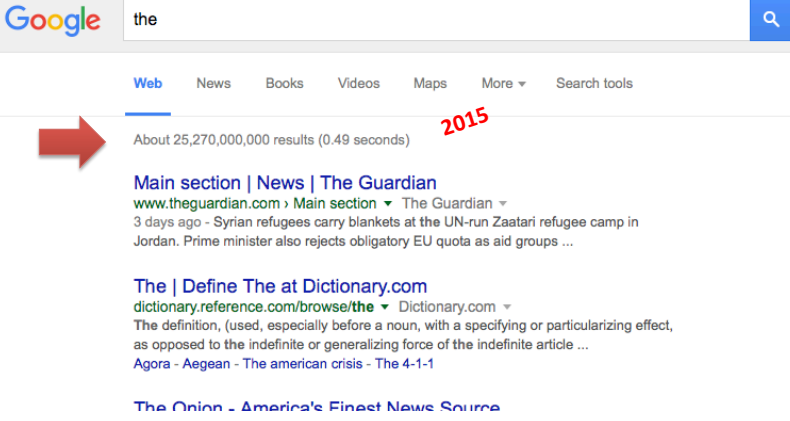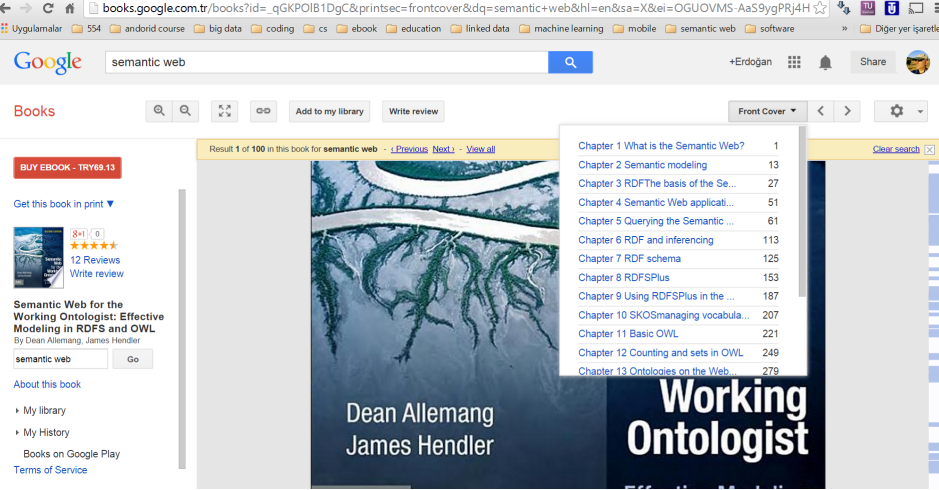
About 9,660,000,000 results (0.27 seconds)   2014

The | Define The at Dictionary.com
dictionary.reference.com/browse/the ▾ Dictionary.com ▾
The definition, (used, especially before a noun, with a specifying or particularizing effect, as opposed to **the** indefinite or generalizing force of **the** indefinite article ...
The adversary - The 4-1-1 - The african roscius - The 411

The Age: Latest & Breaking News Melbourne, Victoria
www.**the**age.com.au/ ▾ The Age ▾
The Age has **the** latest local news on Melbourne, Victoria. Read National News from Australia, World News, Business News and Breaking News stories.

The Onion - America's Finest News Source
www.**the**onion.com/ ▾ The Onion ▾
The Onion, America's Finest News Source, is an award-winning publication covering world, national, and local issues. It is updated daily online and distributed ...

BİL 546 - Semantic Web – TOBB ETÜ
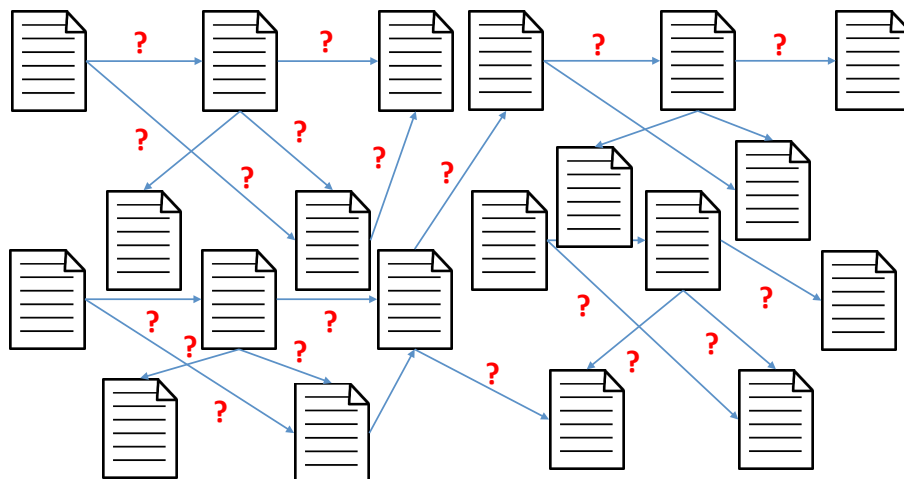
10

How many docs w/«the»?

books.google.com

books.google.com

- How many?
  - >**30 million** books scanned b/w 2004-2013
- How many books in the world?
  - Google estimated around **130 million** books
- When will Google complete scanning?
  - **2020**
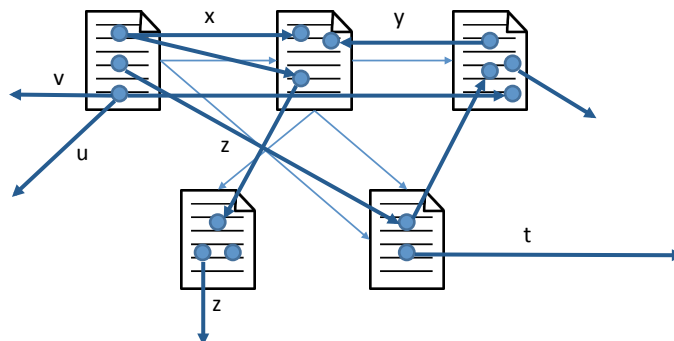
Problem with the web

Problem with the web (cnt.)

- Software cannot use the web automatically and easily,
  - Requires custom parsing, interpretation, …
- Very big to use just by browsing
  - 500 million tweets per day

Web 3.0 (2010)

- Web of Documents (web1.0) to
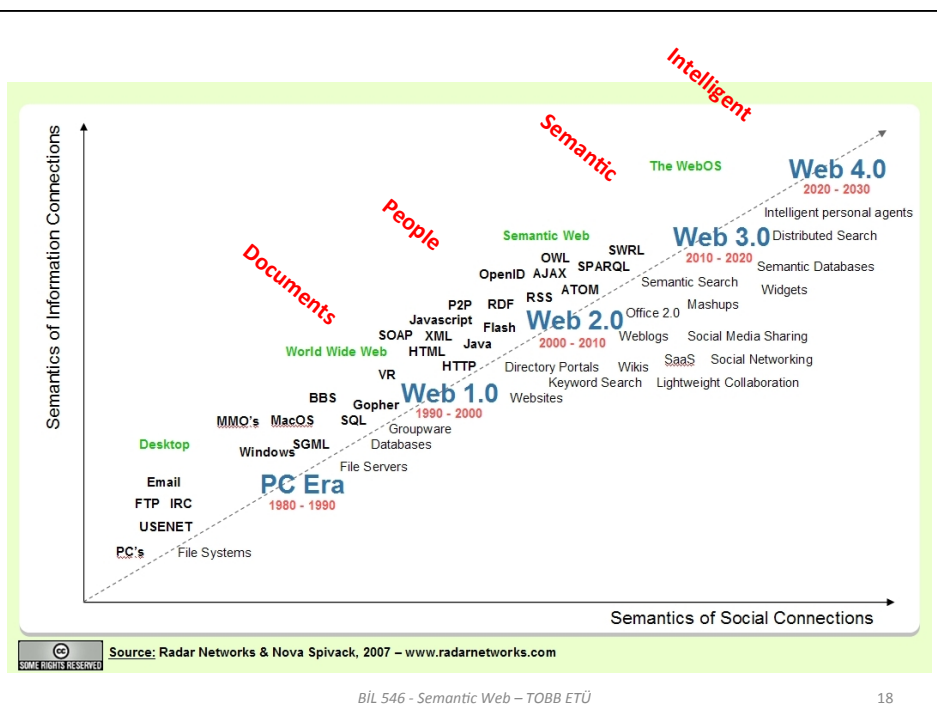- Web of Data

What is Semantic Web?

- "The Semantic Web is **not a separate Web but an extension** of the current one, in which **information is given well-defined meaning**, better **enabling computers and people to work in cooperation**."
  - **Tim-Berners Lee**, Scientific American, May 2001

BİL 546 - Semantic Web – TOBB ETÜ   17

Source: Radar Networks & Nova Spivack, 2007 – www.radarnetworks.com

BİL 546 - Semantic Web – TOBB ETÜ   18

Internet to Web

- Internet *(1960s)*
- Web 1.0 *(1990)*
  - pull web, read-web
- Web 2.0 *(2000s)*
  - pull+push, read-write web, collective content generation
- Web 3.0 *(2010-)*
  - Semantic web
- Web 4.0 (*future*)
  - Intelligent web

BİL 546 - Semantic Web – TOBB ETÜ   19

How to realize semantic web?

- Data modelling
  - Ontologies, ontology languages
- Storage
  - In web docs or in datastores
- Querying
  - SPARQL
- Applications
  - Many

BİL 546 - Semantic Web – TOBB ETÜ   20

What is Ontology?

- "An ontology is a <u>specification</u> of a <u>conceptualization</u>."
  - Tom Gruber

- *Ontologies provide a shared understanding of a domain.*

Ontology

- Conceptualization (of domain knowledge)
  - Classes, relationships, instances, …
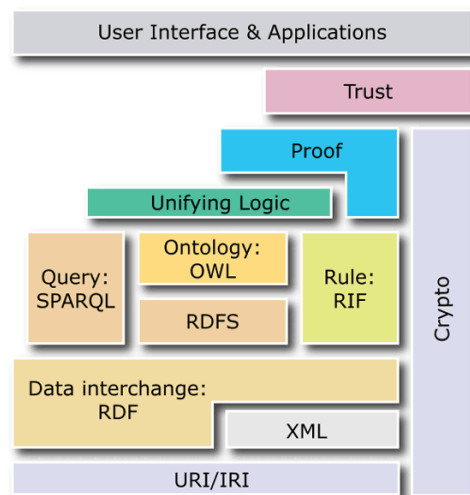- Specification (formally)
  - RDF, RDFS, OWL

W3C

- World Wide Web Consortium (W3C)
  - Int. community defining web stds
  - Founded in 1994 by Tim Berners-Lee
  - 377 member (now)
- Semantic Web standards
  - http://www.w3.org/standards/semanticweb/

Layered Cake / Protocols

- Standards
  - Basic
    - **RDF**
    - RDFS
    - OWL
    - SPARQL
  - More
    - RDFa
    - Schema.org

Semantic Web Vision (W3C)

- Extend principles of the Web **from documents to data**
- Data should be accessed using the general Web architecture (e.g., URI-s, protocols, …)
- Data should be related to one another just as documents are already
- Creation of a common framework that allows:
  - Data to be **shared** and **reused** across applications
  - Data to be **processed automatically**
  - New relationships between pieces of data to be **inferred**

New Vision for the Web

Web of Documents

Web of Data

RDF

- Resource Description Framework

:ankara — :hasArea → "2516 km^2"

property

resource — literal

:capitalOf → :turkey
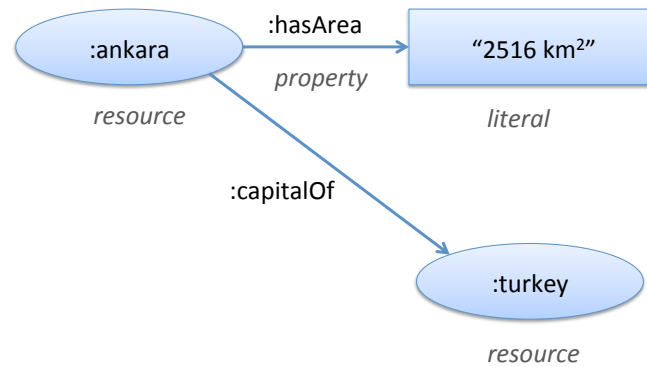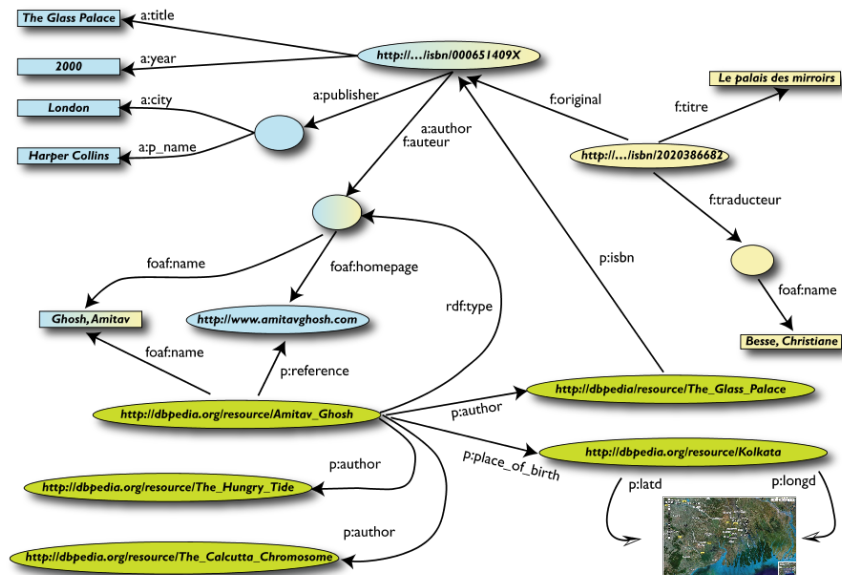
resource

Source: Ivan Herman slides

Semantic Web Tools

- Protege
  - Ontology editor
- Jena
  - Open source framework in Java for app development
- Openlink Virtuoso
  - Storage
- Many more
  - http://www.mkbergman.com/**sweet-tools**/
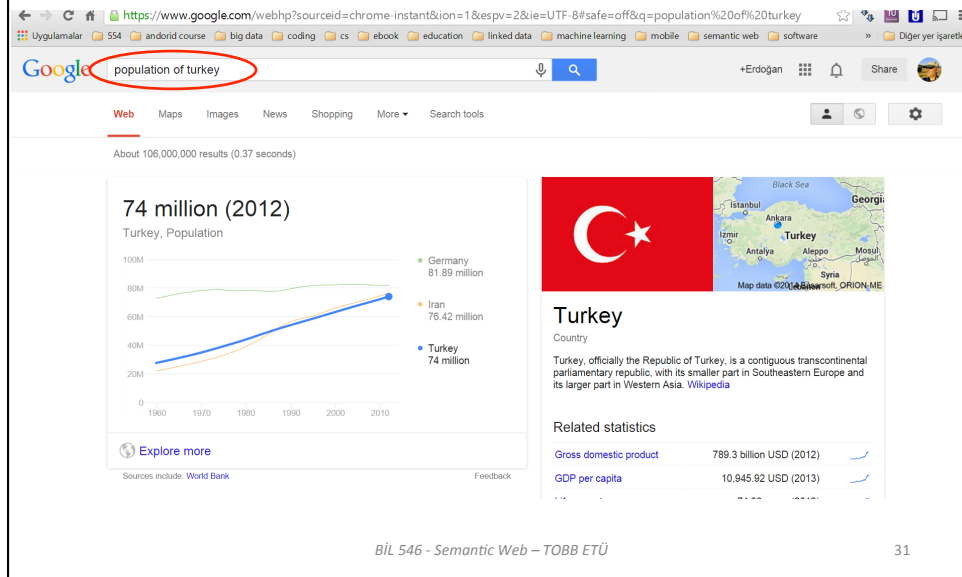
Where is Semantic Web?

- Many applications utilize semantic technologies
  - Apple's Siri
  - BBC (music, sports, etc.)
    - http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/
  - Web search
    - Schema.org, Google Knowledge Graph, Facebook Open Graph Protocol, RDFa, …
  - IBM Watson
  - Linked Data
  - …

Google search

IBM Watson

- An artificially intelligent computer system **capable of answering questions posed in natural language**
- http://www.ibm.com/smarterplanet/us/en/ibmwatson/
- http://en.wikipedia.org/wiki/Watson_(computer)

- IBM's Watson Supercomputer Destroys Humans in Jeopardy
    - http://www.youtube.com/watch?v=WFR3lOm_xhE

Linked Data

- RDF data in big volumes (billions!)
- **Linking Open Data** Project
  - linkeddata.org
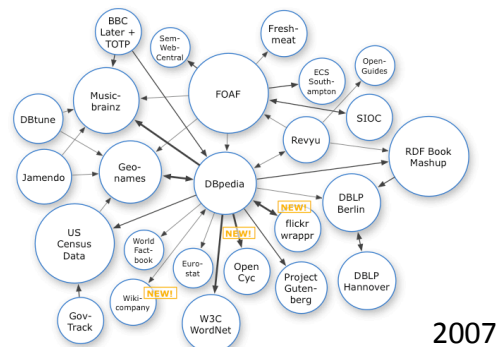
2007

Linked Data
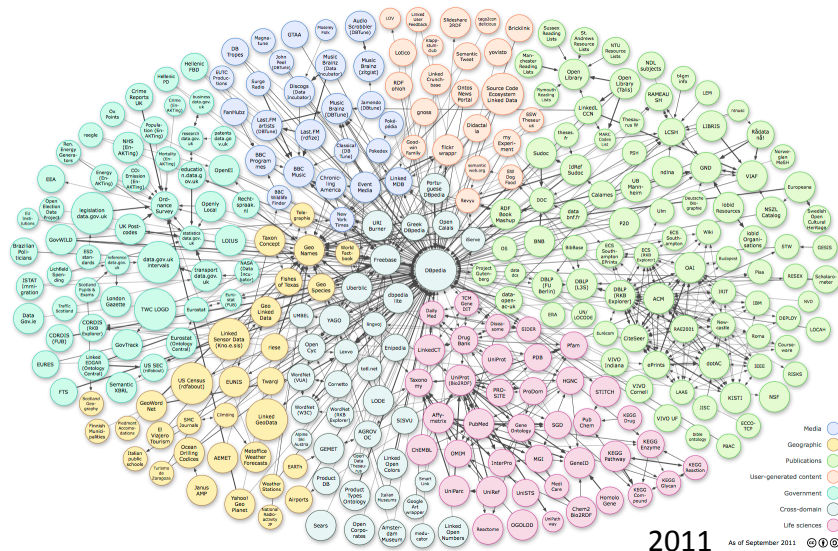

2011   As of September 2011

BİL 546 - Semantic Web – TOBB ETÜ

35

Wikipedia → DBpedia

Infobox
Structured data

BİL 546 - Semantic Web – TOBB ETÜ

36

18

37

DBpedia

- Project started in 2007, by Free Univ. of Berlin, Univ. of Leipzig, OpenLink Software
- Convert all infobox and structured data in wikipedia pages to linkeddata
- Current English dbpedia knowledgebase includes
  - 3.7 million things
    - including 764,000 persons, 573,000 places (including 387,000 populated places), 333,000 creative works (including 112,000 music albums, 72,000 films and 18,000 video games), 192,000 organizations (including 45,000 companies and 42,000 educational institutions), 202,000 species and 5,500 diseases.
- 1.89 billion triples for 111 languages.
- Nucleus for the Web of Data

BİL 546 - Semantic Web – TOBB ETÜ   38

DBpedia

- [List people who were born in Ankara with their name, birthdate, and description (**click** to run)](#)

SPARQL query:

```
SELECT ?name ?birth ?description ?person
WHERE {
    ?person dbpedia2:birthPlace :Ankara .
    ?person dbpedia2:birthDate ?birth .
    ?person foaf:name ?name .
    ?person rdfs:comment ?description .
    FILTER (LANG(?description) = 'en') .
}
```

BİL 546 - Semantic Web – TOBB ETÜ   39

Linked Data

- 295 datasets
- >31 billion triples   *Source: http://www4.wiwiss.fu-berlin.de/lodcloud/state/*

Domain	Number of datasets	Triples	%	(Out-)Links	%
Media	25	1,841,852,061	5.82 %	50,440,705	10.01 %
Geographic	31	6,145,532,484	19.43 %	35,812,328	7.11 %
Government	49	13,315,009,400	42.09 %	19,343,519	3.84 %
Publications	87	2,950,720,693	9.33 %	139,925,218	27.76 %
Cross-domain	41	4,184,635,715	13.23 %	63,183,065	12.54 %
Life sciences	41	3,036,336,004	9.60 %	191,844,090	38.06 %
User-generated content	20	134,127,413	0.42 %	3,449,143	0.68 %
	295	31,634,213,770		503,998,829	

40

Linked Data (2014)

- 1014 datasets

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	51.28%
Total	1014	

Source: http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

41

Books

- **Semantic Web for the Working Ontologist**
  Dean Allemang, Jim Hendler
  2nd ed., Morgan Kaufmann, 2011
- **Practical Semantic Web and Linked Data Applications**
  Mark Watson, 2010
- **A Semantic Web Primer**
  Grigoris Antoniou and Frank van Harmelen
  2nd ed., MIT, 2008

BİL 546 - Semantic Web – TOBB ETÜ   42

Course

- Review standards
  - RDF, RDFS, OWL, SPARQL
- Rules, reasoning, linked data, tools
- 2-3 homeworks (15%)
  - Protege, Jena, Virtuoso
- 1 written exam (35%)
- Project or Paper (50%)

Paper topics

- Semantic search and information retrieval
- Information extraction
- Entity linking
- Question answering
- Document Classification
- …

Paper work

- Review important and recent papers
- Analyse and propose a work plan
- Implement and test
- Write results

Conferences

- **ISWC** International Semantic Web Conference
- **IEEE ICSC** International Conference on Semantic Computing
- **ESWC** Extended Semantic Web Conference
- **WWW** World Wide Web Conference

Journals

- Semantic Web Journal, IOS
- Journal of Web Semantics, Elsevier
- International Journal on Semantic Web and Information Systems, IGI Global

References

- W3C Semantic Web Activity, http://www.w3.org/2001/sw/
- Michael K. Bergman, **A Decade in the Trenches of the Semantic Web** (2014) http://www.mkbergman.com/1771/a-decade-in-the-trenches-of-the-semantic-web/
- http://semanticweb.com/
- http://www.cambridgesemantics.com/

The end