

Introduction to **Semantic Web**

Erdogan Dogdu

TOBB University of Economics & Technology

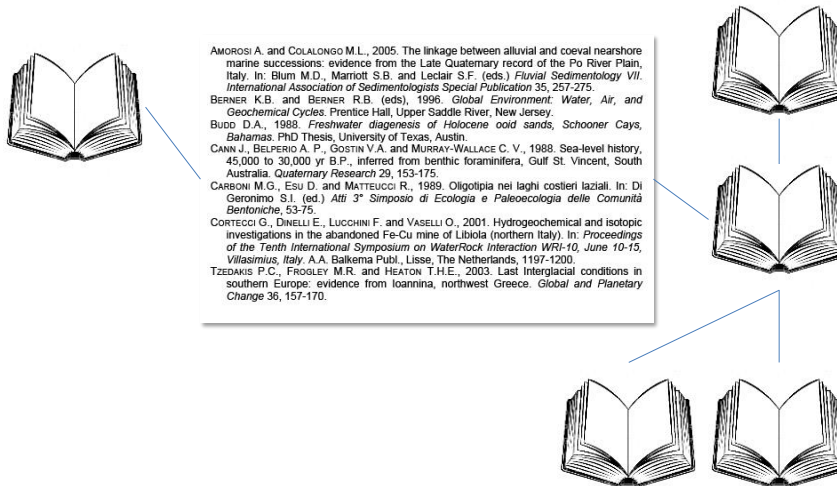
2014



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

Before the Web

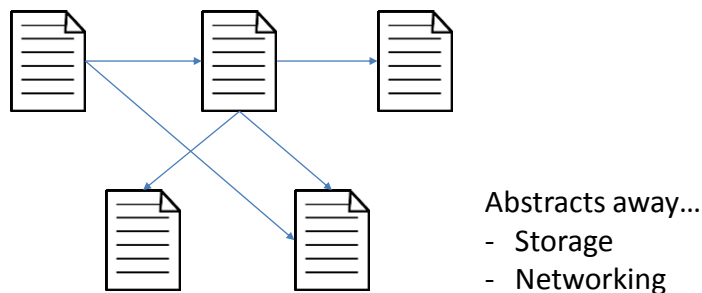


BLI 546 - Semantic Web – TOBB ETÜ

3

Web 1.0

- Hyperlinked web documents



BLI 546 - Semantic Web – TOBB ETÜ

4

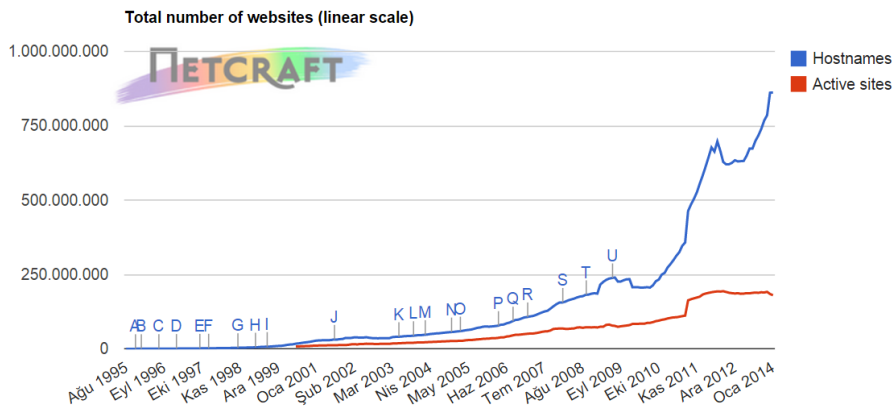
Web 2.0

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

BİL 546 - Semantic Web – TOBB ETÜ

5

Web is big!



BİL 546 - Semantic Web – TOBB ETÜ

6

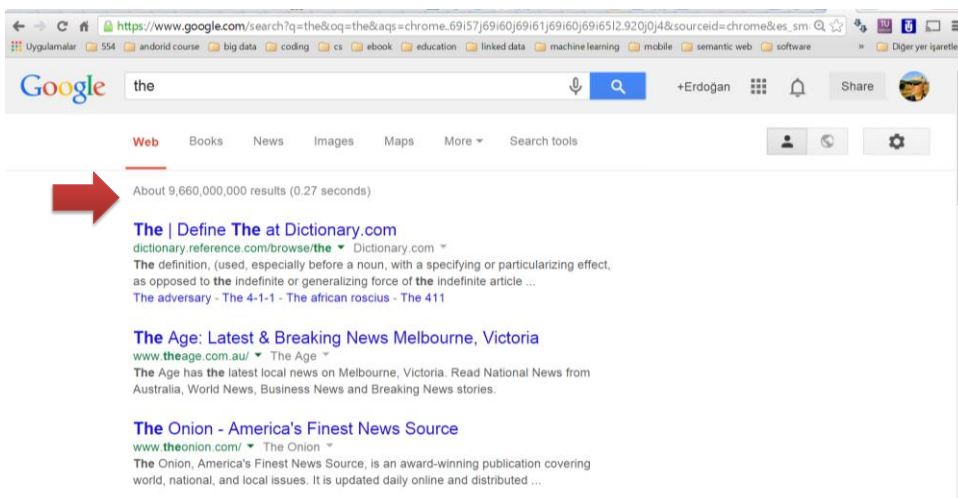
#Web pages in billions...



BİL 546 - Semantic Web – TOBB ETÜ

7

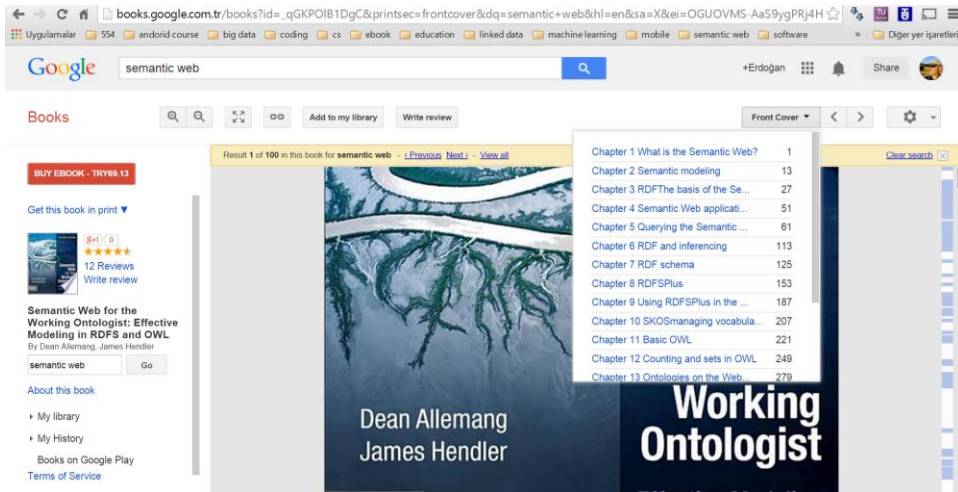
How many docs w/ «the»?



BİL 546 - Semantic Web – TOBB ETÜ

8

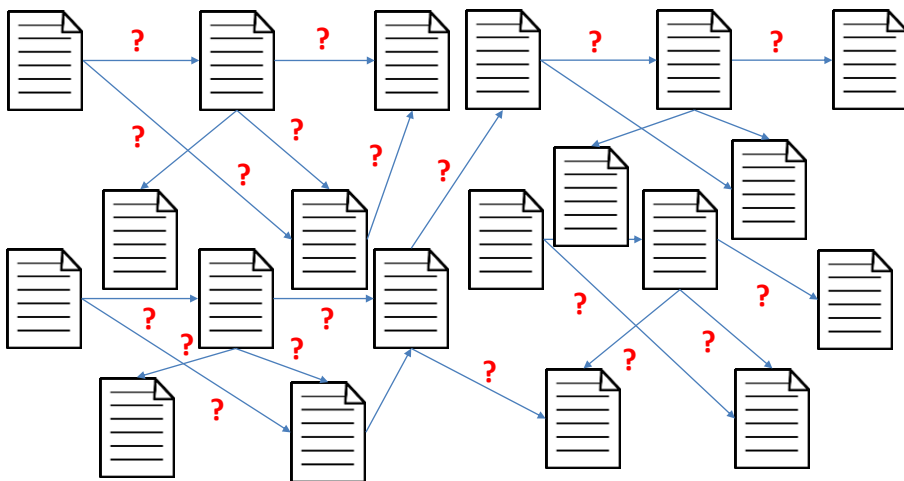
books.google.com



BİL 546 - Semantic Web – TOBB ETÜ

9

Problem with the web



BİL 546 - Semantic Web – TOBB ETÜ

10

Problem with the web (cnt.)

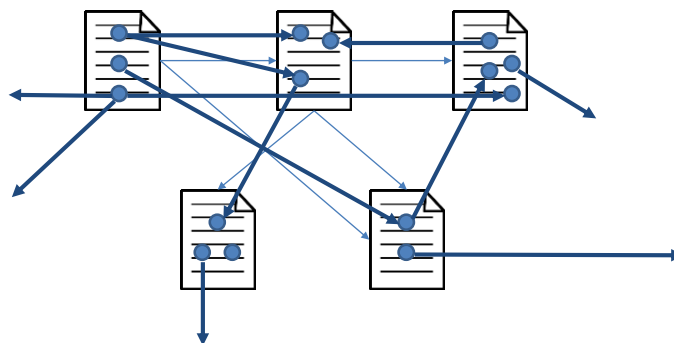
- Software cannot use the web automatically and easily,
 - Requires custom parsing, interpretation, ...
- Too big to use just by browsing
 - 500 million tweets per day

BlL 546 - Semantic Web – TOBB ETÜ

11

Web 3.0

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



BlL 546 - Semantic Web – TOBB ETÜ

12

What is Semantic Web?

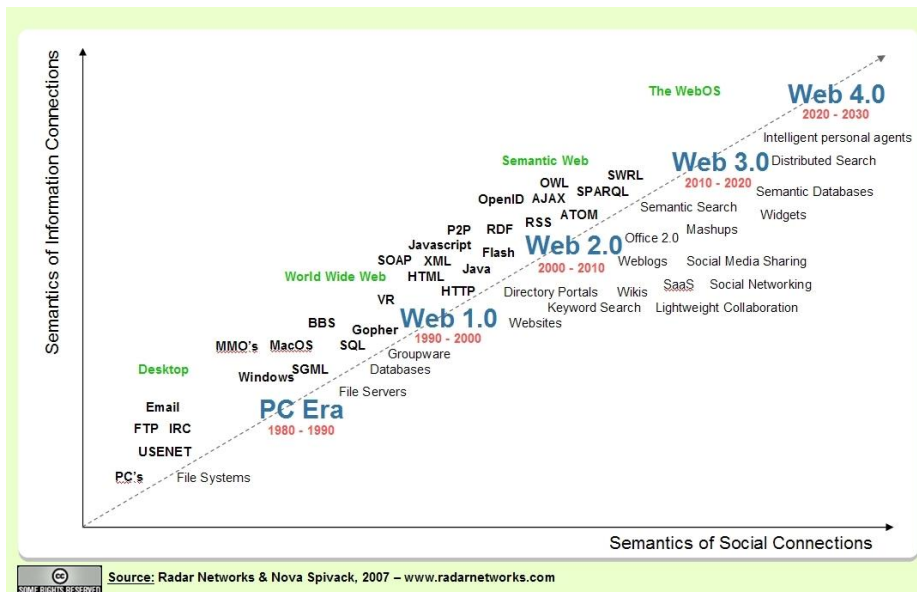
- “The Semantic Web is **not a separate Web but an extension** of the current one, in which **information is given well-defined meaning**, better **enabling computers and people to work in cooperation.**”

- **Tim-Berners Lee**, Scientific American, May 2001



Bil 546 - Semantic Web – TOBB ETÜ

13



Bil 546 - Semantic Web – TOBB ETÜ

14

Internet to Web

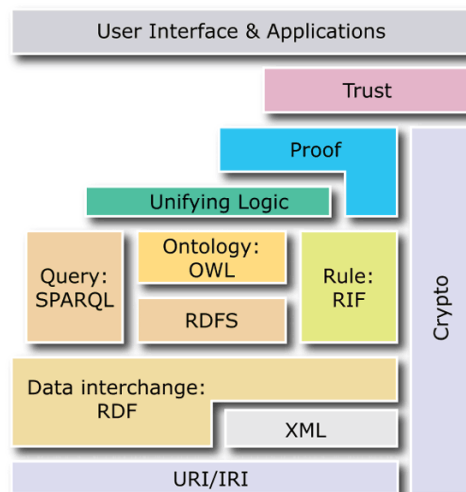
- Internet (1960s)
- Web 1.0 (1990)
 - pull web, read-web
- Web 2.0 (2000s)
 - pull+push, read-write web, collective content generation
- Web 3.0 (future)
 - intelligent web

BlL 546 - Semantic Web – TOBB ETÜ

15

How to realize Semantic Web

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



BlL 546 - Semantic Web – TOBB ETÜ

16

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/>

BlL 546 - Semantic Web – TOBB ETÜ

19

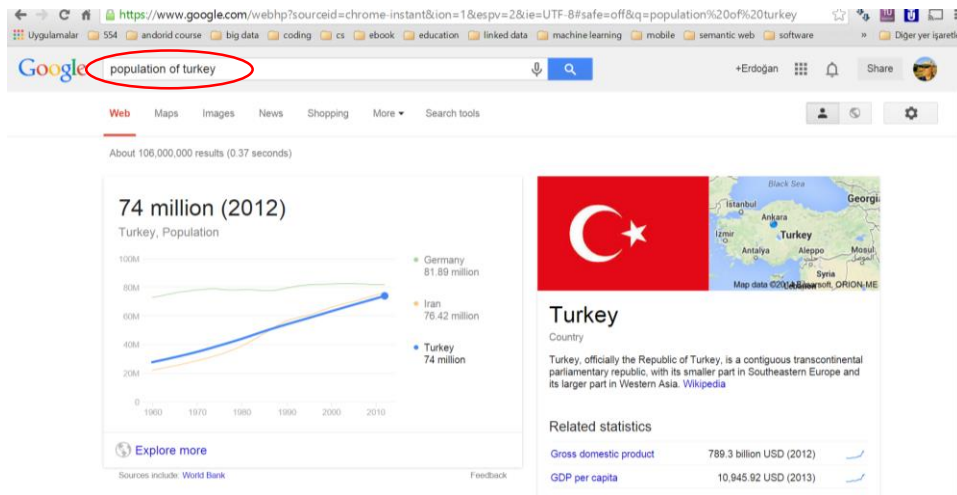
Where is Semantic Web?

- 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
 - ...

BlL 546 - Semantic Web – TOBB ETÜ

20

Google search



BL 546 - Semantic Web – TOBB ETÜ

21

IBM Watson

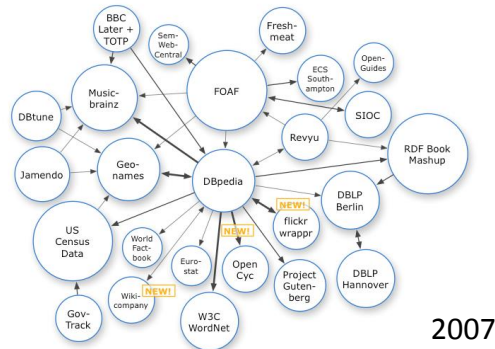
- An artificially intelligent computer system **capable of answering questions posed in natural language**
- <http://www.ibm.com/smarterplanet/us/en/ibmwatson/>
- [http://en.wikipedia.org/wiki/Watson_\(computer\)](http://en.wikipedia.org/wiki/Watson_(computer))
- IBM's Watson Supercomputer Destroys Humans in Jeopardy
 - http://www.youtube.com/watch?v=WFR3IOm_xhE

BL 546 - Semantic Web – TOBB ETÜ

22

Linked Data

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

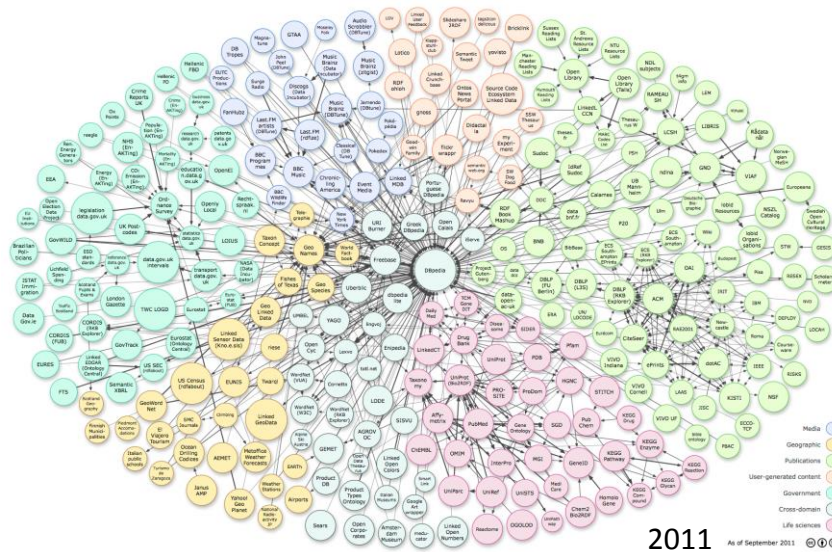


2007

BlL 546 - Semantic Web – TOBB ETÜ

23

Linked Data



2011 As of September 2011

24

Wikipedia → DBpedia

Ankara (Turkish pronunciation: [an.ka.ɾa]; historically known as Angora) is the capital of Turkey and the country's second largest city after Istanbul. The city has a mean elevation of 938 metres (3,077 ft).^[1] Centrally located in Anatolia, Ankara is an important commercial and industrial city. It is the center of the Turkish Government, and houses all foreign embassies. It is an important crossroads of trade, strategically located at the centre of Turkey's highway and railway networks, and serves as the marketing centre for the surrounding agricultural area. The city was famous for its long-haired Angora goat and its prized wool (mohair), a unique breed of cat (Angora cat), Angora rabbits and their prized wool (Angora wool), pears, honey, and the region's muscat grapes.

The historical center of Ankara is situated upon a rocky hill, which rises 150 m (492 ft) above the plain on the left bank of the Ankara Çayı, a tributary of the Sakarya (Sangarius) river. The city is located at 39°52′30″ North, 32°52′ East (39.875°N 32.833°E), about 450 km (280 mi) to the southeast of Istanbul, the country's largest city. Although situated in one of the driest places of Turkey and surrounded mostly by steppe vegetation except for the forested areas on the southern periphery, Ankara can be considered a green city in terms of green areas per inhabitant, which is 72 m² per head.^[2]

Ankara is a very old city with various Hittite, Phrygian, Hellenistic, Roman, Byzantine, and Ottoman archaeological sites. The hill which overlooks the city is crowned by the ruins of the old castle, which adds to the picturesqueness of the view, but only a few historic structures surrounding the old citadel have survived to the present day. There are, however, many finely preserved remains of Hellenistic, Roman and Byzantine architecture, the most remarkable being the Temple of Augustus and Rome (20 BC) which is also known as the *Monumentum Ancyranum*.^[3]

According to the Turkish Statistical Institute, as of 2011 the city of Ankara had a population of 4,338,620 and its metropolitan municipality 4,550,662.^{[4][5]}

Contents [show]

- 1 Etymology
- 2 History
 - 2.1 Ancient history
 - 2.2 Celtic history
 - 2.3 Roman history
 - 2.4 Turkish history
- 3 Climate
- 4 Demographics
- 5 Economy
- 6 Main sights
 - 6.1 Museums
 - 6.2 Archaeological sites
 - 6.2.1 Ankara Citadel

Infobox
Structured data

BİL 546 - Semantic Web – TOBB ETÜ

25

About: Ankara

An Entity of Type: [List of cities in Turkey](#), from Named Graph: <http://live.dbpedia.org>, within Data Space: live.dbpedia.org

Ankara (historically known as Angora) is the capital of Turkey and the country's second largest city after Istanbul. The city has a mean elevation of 938 metres (3,077 ft). Centrally located in Anatolia, Ankara is an important commercial and industrial city. It is the center of the Turkish Government, and houses all foreign embassies.

Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	2516.0
dbpedia-owl:PopulatedPlace/populationDensity	1551.0
dbpedia-owl:abstract	Ankara (historically known as Angora) is the capital of Turkey and the country's second largest city after Istanbul. The city has a mean elevation of 938 metres (3,077 ft). Centrally located in Anatolia, Ankara is an important commercial and industrial city. It is the center of the Turkish Government, and houses all foreign embassies. It is an important crossroads of trade, strategically located at the centre of Turkey's highway and railway network, and serves as the marketing centre for the surrounding agricultural area. The city was famous for its long-haired Angora goat and its prized wool, a unique breed of cat (Angora cat), Angora rabbits and their prized wool (Angora wool), pears, honey, and the region's muscat grapes. The historical center of Ankara is situated upon a rocky hill, which rises 150 m (492 ft) above the plain on the left bank of the Ankara Çayı, a tributary of the Sakarya (Sangarius) river. The city is located at 39°52′30″ North, 32°52′ East (39.875°N 32.833°E), about 450 km (280 mi) to the southeast of Istanbul, the country's largest city. Although situated in one of the driest places of Turkey and surrounded mostly by steppe vegetation except for the forested areas on the southern periphery, Ankara can be considered a green city in terms of green areas per inhabitant, which is 72 m² per head. Ankara is a very old city with various Hittite, Phrygian, Hellenistic, Roman, Byzantine, and Ottoman archaeological sites. The hill which overlooks the city is crowned by the ruins of the old castle, which adds to the picturesqueness of the view, but only a few historic structures surrounding the old citadel have survived to the present day. There are, however, many finely preserved remains of Hellenistic, Roman and Byzantine architecture, the most remarkable being the Temple of Augustus and Rome (20 BC) which is also known as the <i>Monumentum Ancyranum</i>. According to the Turkish Statistical Institute, as of 2011 the city of Ankara had a population of 4,338,620 and its metropolitan municipality 4,550,662.
dbpedia-owl:areaCode	312
dbpedia-owl:areaTotal	2516000000.000000 (xsd:double)
dbpedia-owl:country	dbpedia:Turkey
dbpedia-owl:elevation	938.000000 (xsd:double)
dbpedia-owl:isPartOf	dbpedia:Ankara_Province dbpedia:Central_Anatolia_Region
dbpedia-owl:leaderName	dbpedia:Melih_Gökçek

26

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

BlL 546 - Semantic Web – TOBB ETÜ

27

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') .
}
```

BlL 546 - Semantic Web – TOBB ETÜ

28

Linked Data

- 295 datasets
- >31 billion triples Source: <http://www4.wiwiiss.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	

29

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

Course

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

BI.L 546 - Semantic Web – TOBB ETÜ

31

Paper topics

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

BI.L 546 - Semantic Web – TOBB ETÜ

32

Paper work

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

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/>