

Introduction to Semantic Web

Erdogan Dogdu



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

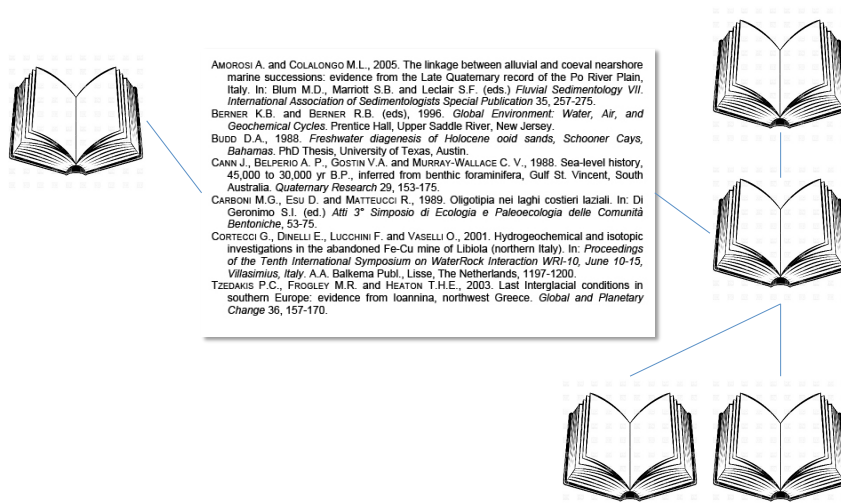
Semantic web

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

Intro to Semantic Web – Erdogan Dogdu

3

Before the Web



Intro to Semantic Web – Erdogan Dogdu

4

Memex

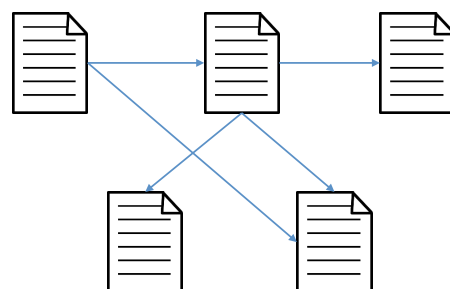
- "As We May Think", Vannevar Bush, 1945, *The Atlantic Monthly*
- **Mem(ory-Ind)ex**
- All information (books, communication, etc) in the form of microfilms and indexed using electromechanical device.

Intro to Semantic Web – Erdogan Dogdu

5

Web 1.0 (1990)

- Hyperlinked web documents



Abstracts away...

- Storage
- Networking

Intro to Semantic Web – Erdogan Dogdu

6

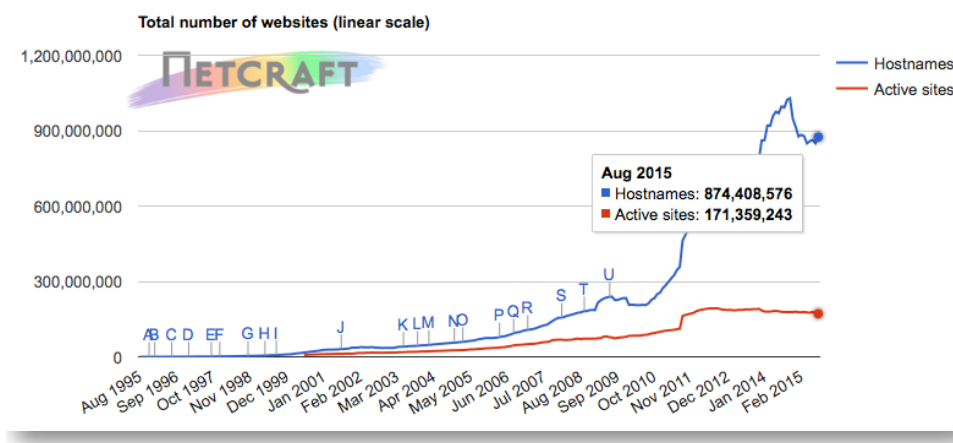
Web 2.0 (2000)

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

Intro to Semantic Web – Erdogan Dogdu

7

Web is big!



Intro to Semantic Web – Erdogan Dogdu

8

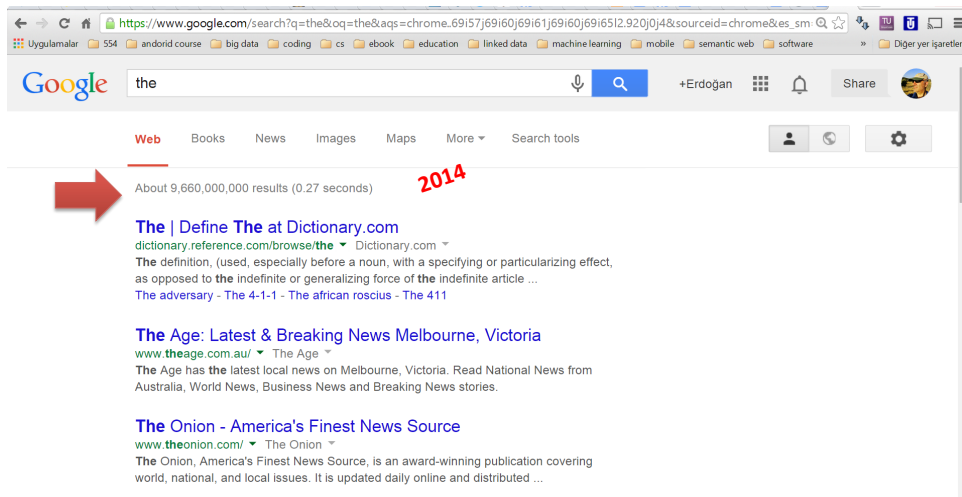
#Web pages in billions...



Intro to Semantic Web – Erdogan Dogdu

9

How many docs w/ «the»?



Intro to Semantic Web – Erdogan Dogdu

10

How many docs w/ «the»?



Google the

Web News Books Videos Maps More Search tools

About 25,270,000,000 results (0.49 seconds) **2015**

Main section | News | The Guardian
www.theguardian.com › Main section ▾ The Guardian ▾
 3 days ago - Syrian refugees carry blankets at the UN-run Zaatari refugee camp in Jordan. Prime minister also rejects obligatory EU quota as aid groups ...

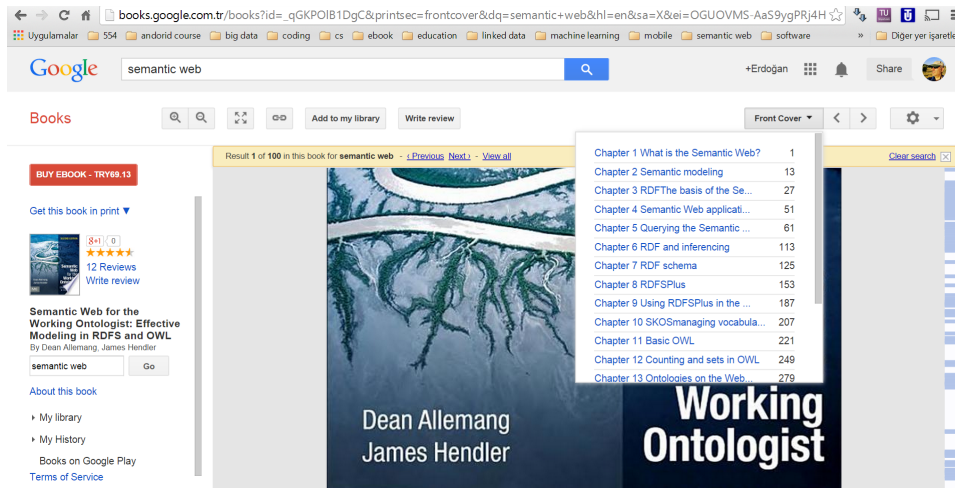
The | Define The at Dictionary.com
dictionary.reference.com/browse/the ▾ Dictionary.com ▾
 The definition, (used, especially before a noun, with a specifying or particularizing effect, as opposed to the indefinite or generalizing force of the indefinite article ...
 Agora - Aegean - The american crisis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

The Onion - America's Finest News Source

Intro to Semantic Web – Erdogan Dogdu

11

books.google.com



books.google.com.tr/books?id=_qGKPOlB1DgC&printsec=frontcover&dq=semantic+web&hl=en&sa=X&ei=OGUOVMS-Aa59ygPRj4H

semantic web

Books

BUY EBOOK - TRY €9.13

Get this book in print ▾

Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL
 By Dean Allemang, James Hendler

semantic web Go

About this book

- My library
- My History
- Books on Google Play
- Terms of Service

Result 1 of 100 in this book for semantic web - [Previous](#) [Next](#) - [View all](#)

Dean Allemang
James Hendler

Working Ontologist

Chapter 1 What is the Semantic Web?	1
Chapter 2 Semantic modeling	13
Chapter 3 RDFThe basis of the Se...	27
Chapter 4 Semantic Web applicati...	51
Chapter 5 Querying the Semantic ...	61
Chapter 6 RDF and inferencing	113
Chapter 7 RDF schema	125
Chapter 8 RDFSPlus	153
Chapter 9 Using RDFSPlus in the ...	187
Chapter 10 SKOSmanaging vocabula...	207
Chapter 11 Basic OWL	221
Chapter 12 Counting and sets in OWL	249
Chapter 13 Ontologies on the Web	279

Intro to Semantic Web – Erdogan Dogdu

12

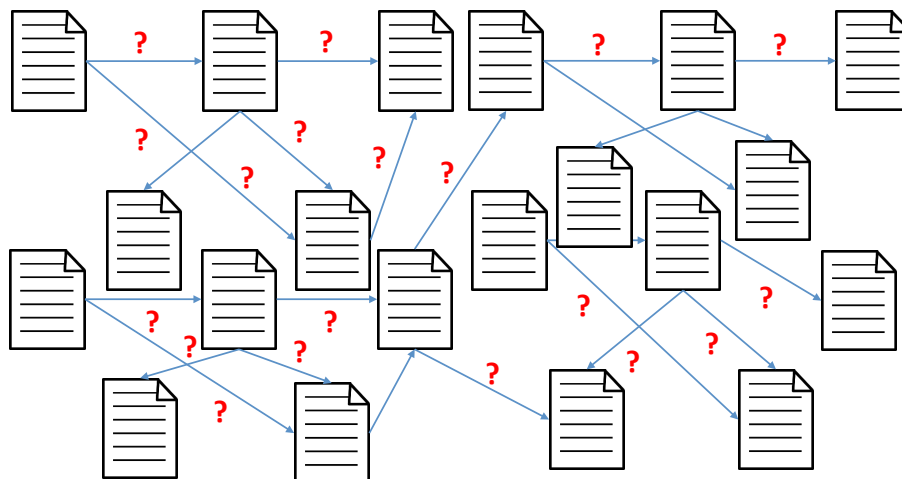
books.google.com

- 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**

Intro to Semantic Web – Erdogan Dogdu

13

Problem with the web



Intro to Semantic Web – Erdogan Dogdu

14

Problem with the web (cnt.)

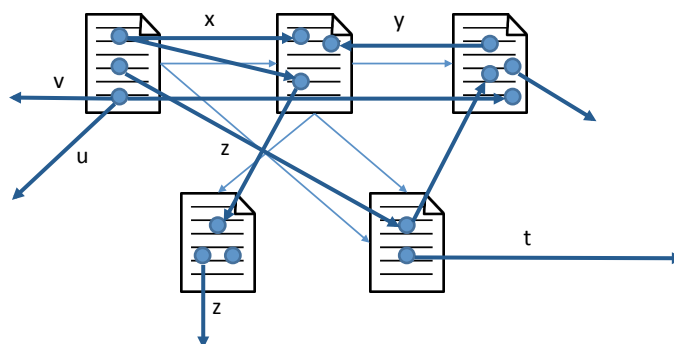
- Software cannot use the web automatically and easily,
 - Requires custom parsing, interpretation, ...
- Very big to use just by browsing
 - 500 million tweets per day

Intro to Semantic Web – Erdogan Dogdu

15

Web 3.0 (2010)

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



Intro to Semantic Web – Erdogan Dogdu

16

What is Semantic Web?

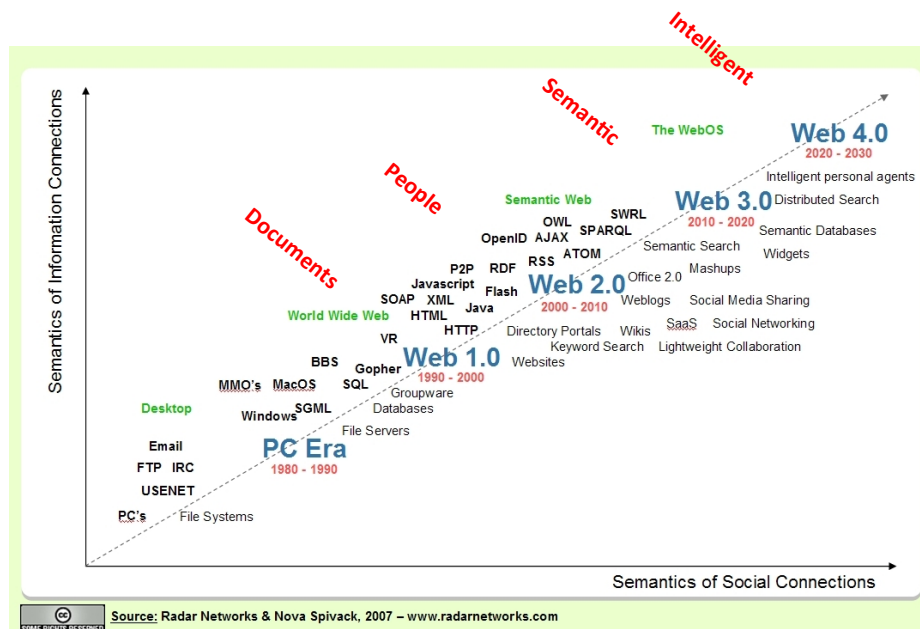
- “The Semantic Web is **not a separate Web but an extension** of the current one, in which **information is given well-defined meaning**, better **enabling computers and people to work in cooperation**.”

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



Intro to Semantic Web – Erdogan Dogdu

17



Intro to Semantic Web – Erdogan Dogdu

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

Intro to Semantic Web – Erdogan Dogdu

19

How to realize semantic web?

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

Intro to Semantic Web – Erdogan Dogdu

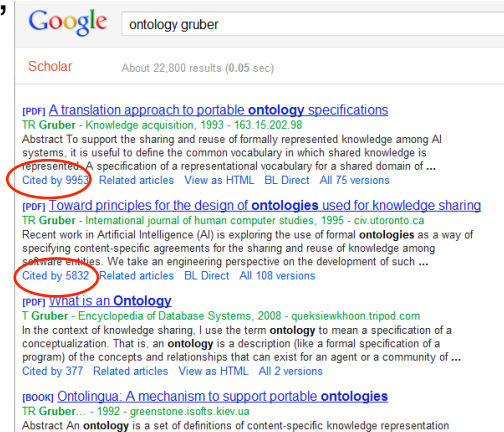
20

What is Ontology?

- “An ontology is a specification of a conceptualization.”

- Tom Gruber

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



Intro to Semantic Web – Erdogan Dogdu

21

Ontology

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

Intro to Semantic Web – Erdogan Dogdu

22

W3C

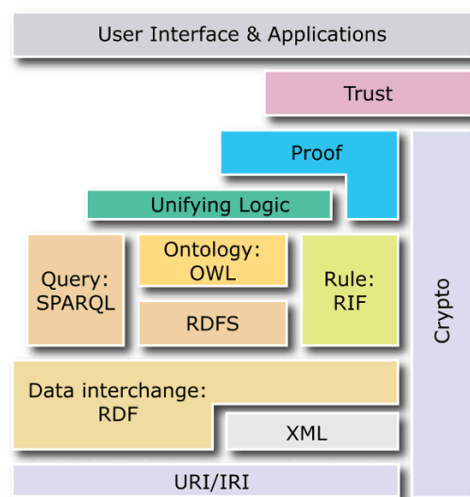
- World Wide Web Consortium (W3C)
 - Int. community defining web stds
 - Founded in 1994 by Tim Berners-Lee
 - 377 member (now)
- Semantic Web standards
 - <http://www.w3.org/standards/semanticweb/>

Intro to Semantic Web – Erdogan Dogdu

23

Layered Cake / Protocols

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



Intro to Semantic Web – Erdogan Dogdu

24

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**

Intro to Semantic Web – Erdogan Dogdu

25

New Vision for the Web

Web of Documents



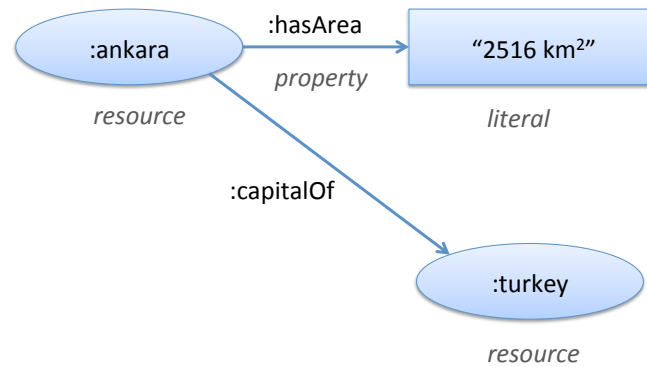
Web of Data

Intro to Semantic Web – Erdogan Dogdu

26

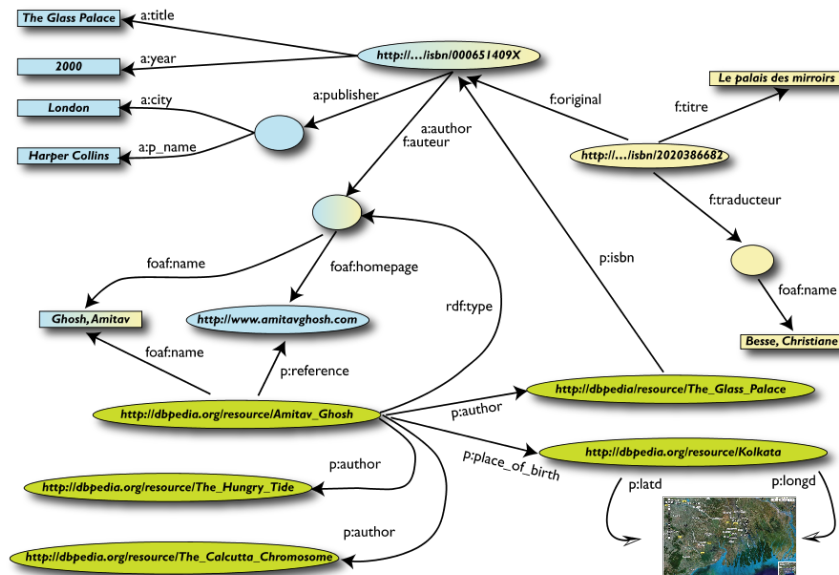
RDF

- Resource Description Framework



Intro to Semantic Web – Erdogan Dogdu

27



Source: Ivan Herman slides

Intro to Semantic Web – Erdogan Dogdu

28

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/>

Intro to Semantic Web – Erdogan Dogdu

29

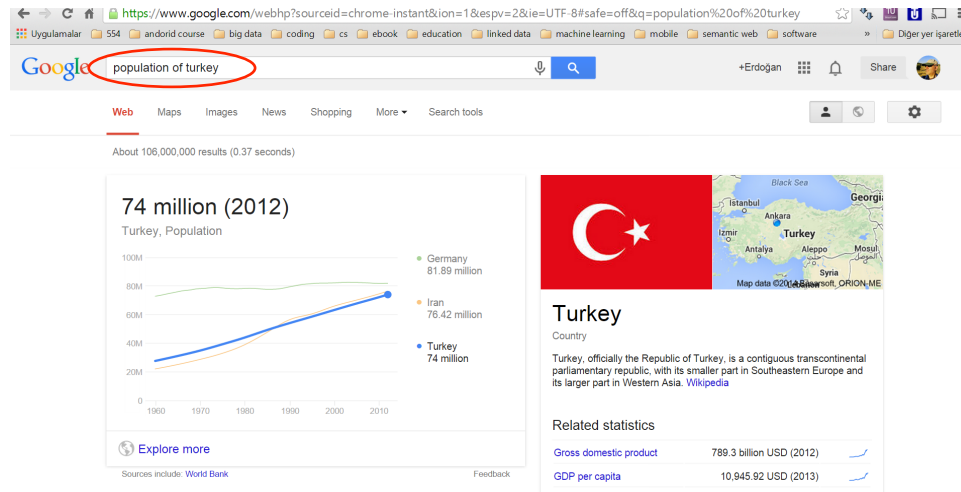
Where is Semantic Web?

- 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
 - ...

Intro to Semantic Web – Erdogan Dogdu

30

Google search



Intro to Semantic Web – Erdogan Dogdu

31

IBM Watson

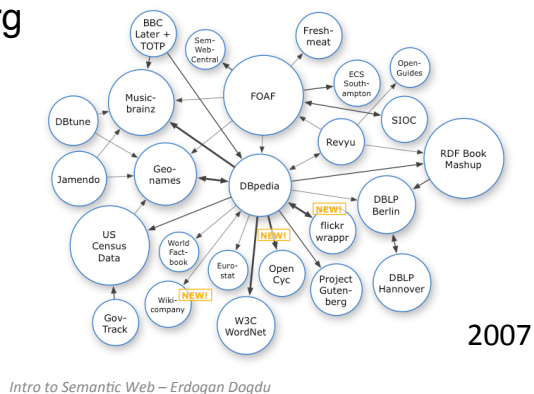
- An artificially intelligent computer system **capable of answering questions posed in natural language**
- <http://www.ibm.com/smarterplanet/us/en/ibmwatson/>
- [http://en.wikipedia.org/wiki/Watson_\(computer\)](http://en.wikipedia.org/wiki/Watson_(computer))
- IBM's Watson Supercomputer Destroys Humans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

Intro to Semantic Web – Erdogan Dogdu

32

Linked Data

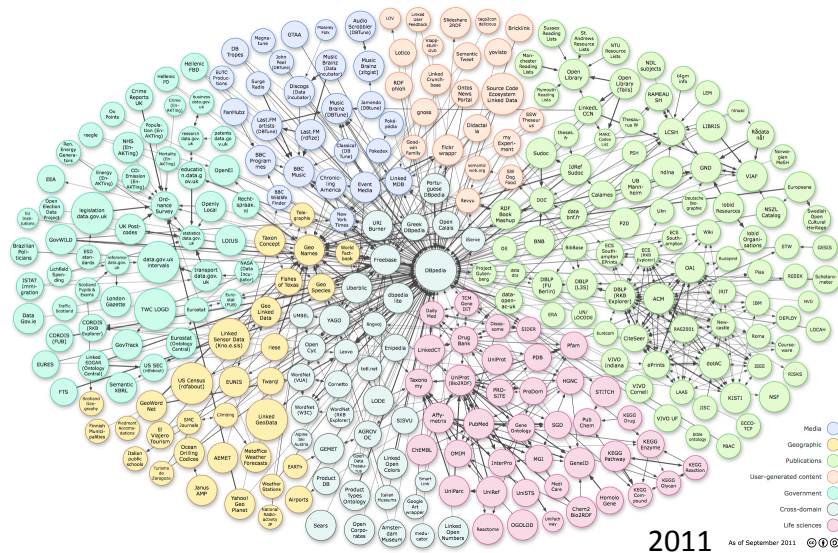
- RDF data in big volumes (billions!)
- **Linking Open Data Project**
 - linkeddata.org



Intro to Semantic Web – Erdogan Dogdu

33

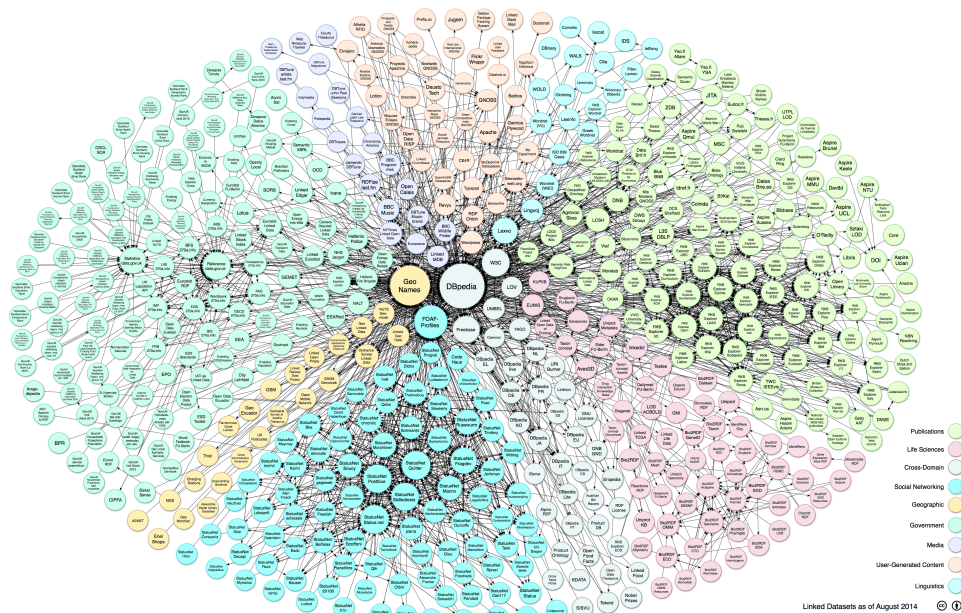
Linked Data



As of September 2011

Intro to Semantic Web – Erdogan Dogdu

34



Intro to Semantic Web – Erdogan Dogdu

35

Wikipedia → DBpedia

The screenshot shows the Wikipedia article for Ankara. The infobox is highlighted with a red box and labeled 'Infobox Structured data'. The infobox contains the following information:

Ankara	City
	
	
Coordinates	39°52′N 32°53′E﻿ / ﻿39.867°N 32.883°E﻿ / 39.867; 32.883
Country	 Turkey
Region	Central Anatolia
Province	Ankara
Government	
 • Mayor	Metin Çoban (AKP)
 • Governor	Abdullah Türel
Area	
 • City	2,516 km ² (971 sq mi)
Elevation	838 m (2,751 ft)

Infobox
Structured data

Intro to Semantic Web – Erdogan Dogdu

36

[illegible]

Intro to Semantic Web – Erdogan Dogdu

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

Intro to Semantic Web – Erdogan Dogdu

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') .
}
```

Intro to Semantic Web – Erdogan Dogdu

39

Linked Data

- 295 datasets
- >31 billion triples

[Source: http://www4.wiwiiss.fu-berlin.de/lodcloud/state/](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	

Intro to Semantic Web – Erdogan Dogdu

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/>

Intro to Semantic Web – Erdogan Dogdu

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

Intro to Semantic Web – Erdogan Dogdu

42

Course

- Review standards
 - www.w3.org/standards/semanticweb
 - RDF, RDFS, OWL, SPARQL
- Rules, reasoning, linked data, tools
- 2-3 homeworks (15%)
 - Protege, Jena, Virtuoso
- 1 written exam (35%)
- Project or Paper (50%)

Intro to Semantic Web – Erdogan Dogdu

43

Paper topics

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

Intro to Semantic Web – Erdogan Dogdu

44

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