

# Introduction to **Semantic Web**

**Erdogan Dogdu**

TOBB University of Economics & Technology

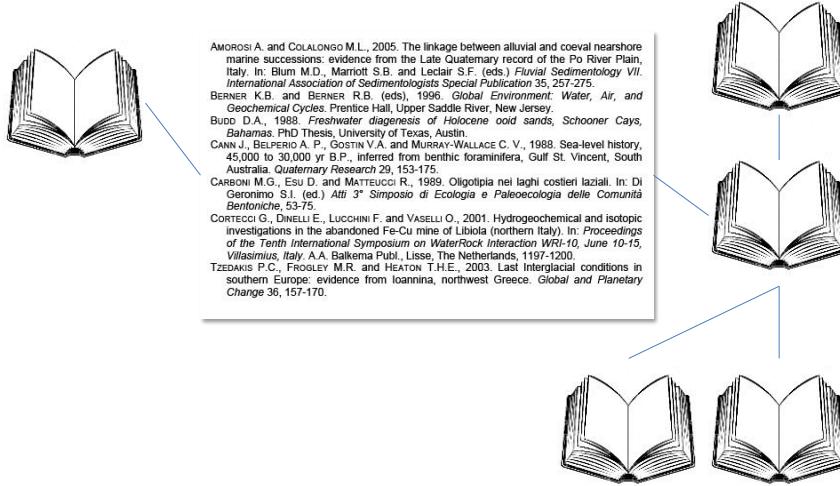
2014



## Outline

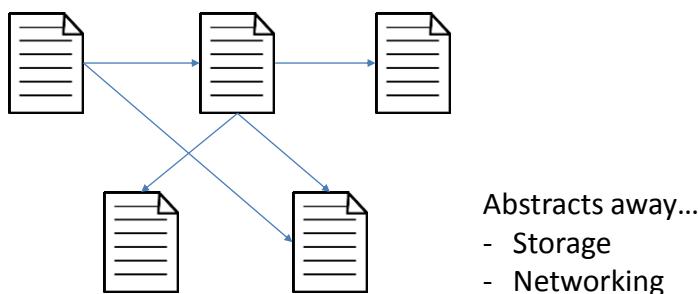
- Motivation
  - Development of the Web: Web 1.0, 2.0, 3.0
  - Limitations of the current Web
- Technical solution
  - Introduction to Semantic Web
  - Semantic Web – architecture and languages
  - Semantic Web - data
  - Semantic Web – tools, processes
- Recent trends
- Summary
- References

# Before the Web



# Web 1.0

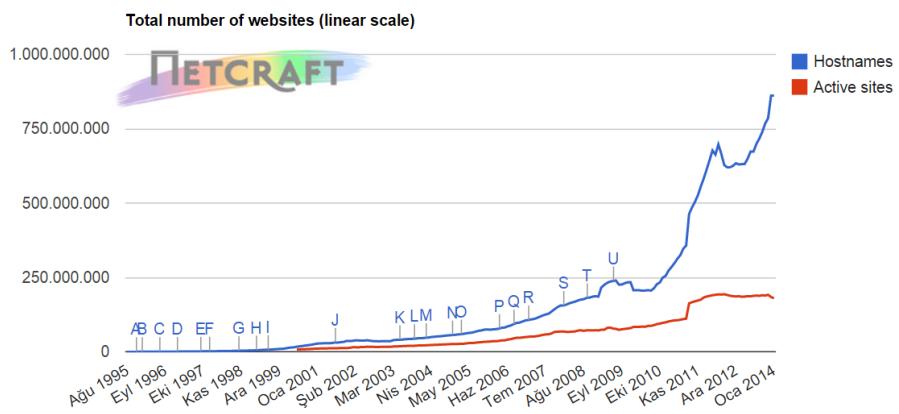
- Hyperlinked web documents



# Web 2.0

- 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...

The screenshot shows the homepage of WorldWideWebSize.com. At the top, it displays the site's name and a subtitle: "DAILY ESTIMATED SIZE OF THE WORLD WIDE WEB". Below this, a main heading reads "The size of the World Wide Web (The Internet)". A summary statement says, "The Indexed Web contains at least 3.32 billion pages (Thursday, 10 July, 2014)." Another statement below it says, "The Dutch Indexed Web contains at least 244.19 million pages (Thursday, 10 July, 2014)." To the right of these statements are social sharing icons for Google+, Facebook, Twitter, and a "Share" button. Below the summary text is a chart titled "The size of the indexed World Wide Web (Number of webpages)". The chart is a line graph showing the number of webpages over time. The y-axis ranges from 45 to 50. The x-axis shows time periods: Last Month, Last Three Months, Last Year, and Last Two Years. The line starts at approximately 45.5, dips slightly, then rises to about 46.5, stays flat, then rises again to about 47.5. Below the chart, there are four buttons: "Last Month", "Last Three Months", "Last Year", and "Last Two Years".

BİL 546 - Semantic Web – TOBB ETÜ

7

## How many docs w/ «the»?

The screenshot shows a Google search results page for the query "the". A large red arrow points to the search bar where "the" is typed. Above the search bar, the URL is https://www.google.com/search?q=the&oq=the&aqs=chrome\_69i57j69i60j69i61j69i60j69i65l2.920j0j4&sourceid=chrome&es\_sm=93. To the right of the search bar are various Google interface elements: a microphone icon, a magnifying glass icon, a user profile icon, a share icon, and a settings gear icon. Below the search bar, the word "Web" is highlighted in red, indicating the search type. The search results begin with a link to Dictionary.com, followed by links to The Age news website and The Onion news source.

BİL 546 - Semantic Web – TOBB ETÜ

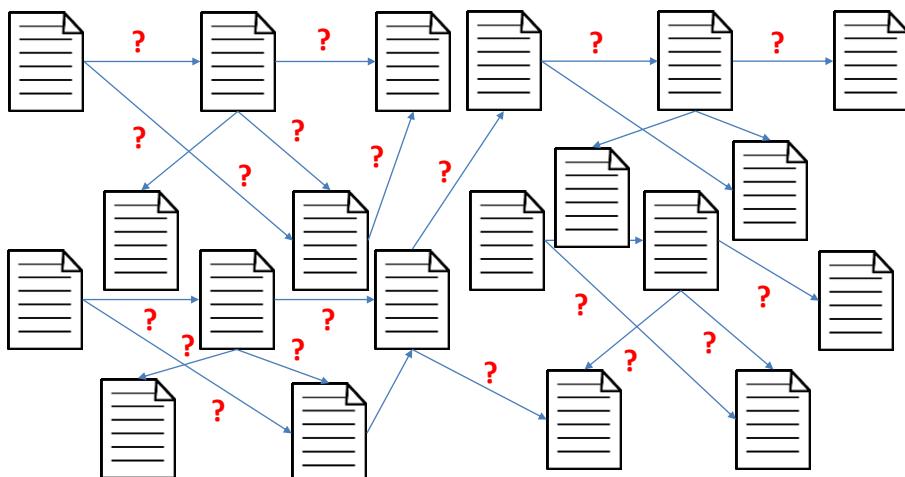
8

# books.google.com

The screenshot shows a Google Books search result for the query "semantic web". The main search bar has "semantic web" typed into it. Below the search bar, there are buttons for "Books", "Search", and "Advanced search". To the right of the search bar, there are links for "Front Cover", "Share", and "Settings". The search results list "Result 1 of 100 in this book for semantic web" followed by "12 Reviews" and a "Write review" button. On the left, there's a sidebar with a "BUY EBOOK - TRY 68 13" button, a "Get this book in print" link, a star rating of 4.5 stars from 12 reviews, and a "Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL" book cover thumbnail. The main content area displays the front cover of the book "Working Ontologist" by Dean Allemang and James Hendler. The cover features a blue background with green foliage at the bottom. The title "Working Ontologist" is prominently displayed in white. To the right of the cover, a table of contents is visible:

Chapter	Page
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 applicat...	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

## Problem with the web

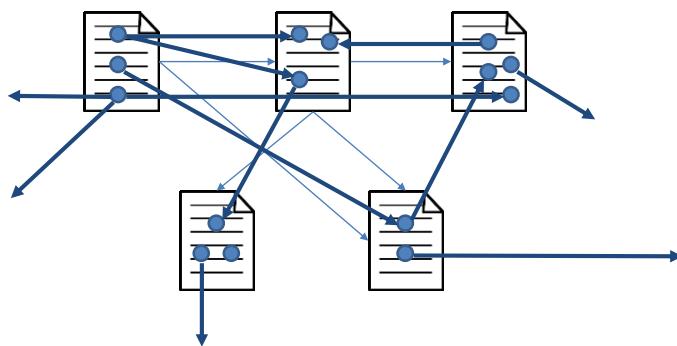


## Problem with the web (cnt.)

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

## Web 3.0

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



# What is Semantic Web?

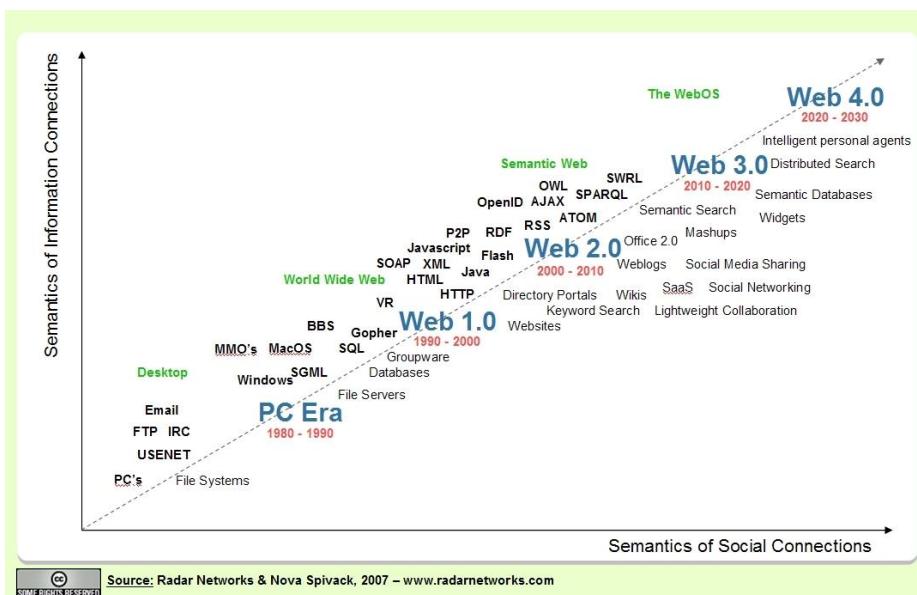
- “The Semantic Web is **not a separate Web but an extension** of the current one, in which **information is given well-defined meaning**, better enabling computers and people to work in cooperation.”

- Tim-Berners Lee, Scientific American, May 2001



BİL 546 - Semantic Web – TOBB ETÜ

13



BİL 546 - Semantic Web – TOBB ETÜ

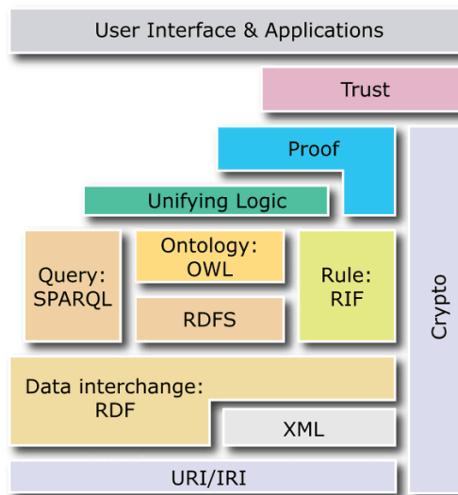
14

# 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

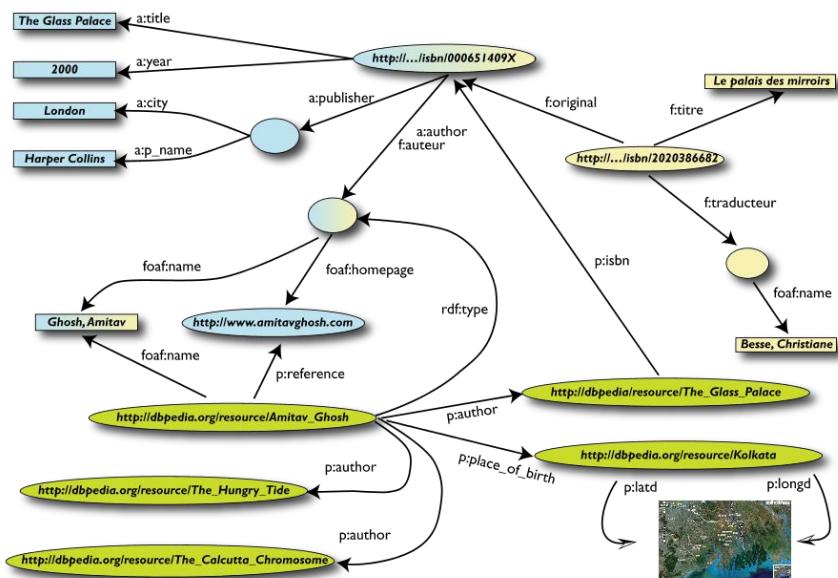
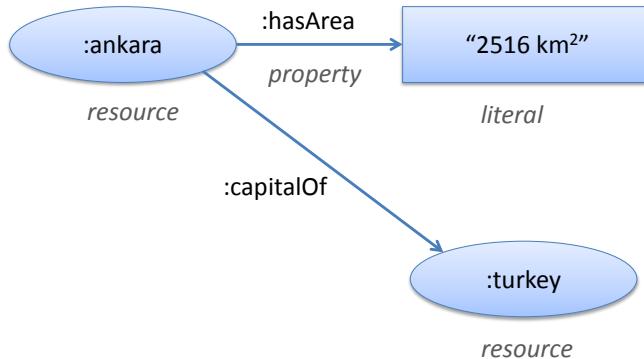
# How to realize Semantic Web

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



# RDF

- Resource Description Framework



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/swo/public/UseCases/BBC/>
  - Web search
    - Schema.org, Google Knowledge Graph, Facebook Open Graph Protocol, RDFa, ...
  - IBM Watson
  - Linked Data
  - ...

# Google search

**74 million (2012)**

Turkey, Population

Year	Germany	Iran	Turkey
1960	~60M	~20M	~20M
1970	~65M	~25M	~25M
1980	~70M	~30M	~30M
1990	~75M	~35M	~35M
2000	~80M	~40M	~40M
2010	~85M	~50M	~74 million

**Turkey**

Country

Turkey, officially the Republic of Turkey, is a contiguous transcontinental parliamentary republic, with its smaller part in Southeastern Europe and its larger part in Western Asia. [Wikipedia](#)

**Related statistics**

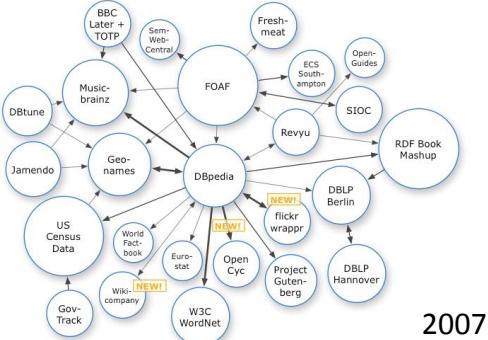
Statistic	Value
Gross domestic product	789.3 billion USD (2012)
GDP per capita	10,945.92 USD (2013)

# 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=WFR3lOm\\_xhE](http://www.youtube.com/watch?v=WFR3lOm_xhE)

# Linked Data

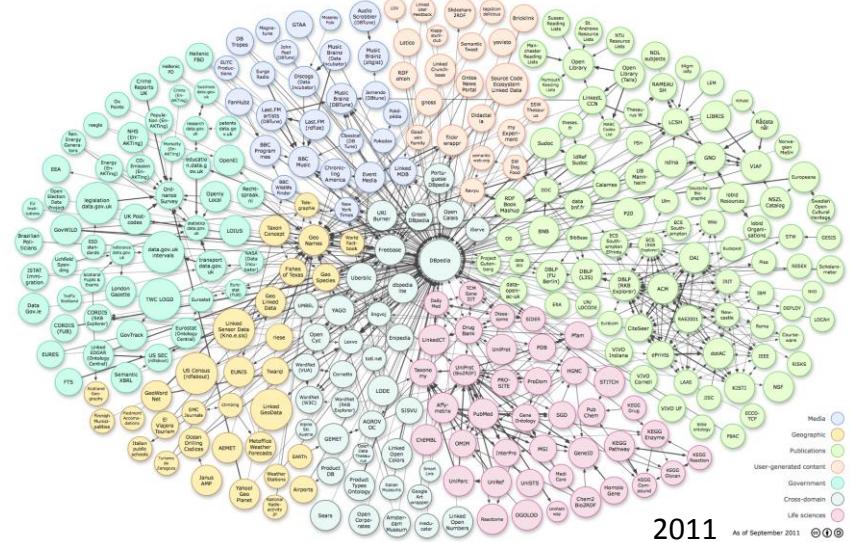
- RDF data in big volumes (billions!)
- **Linking Open Data Project**
  - linkeddata.org



BIL 546 - Semantic Web – TOBB ETÜ

23

# Linked Data



24

# Wikipedia → DBpedia

en.wikipedia.org/wiki/Ankara

**Infobox**  
Structured data

Ankara (Turkish pronunciation: [aŋˈkaɾa]; 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).<sup>20</sup> 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 cultural center, 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). Ankara is also known for 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°42' 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 steep periphery, Ankara can be considered a green city in terms of green areas per inhabitant, which is 72 m<sup>2</sup> per head.<sup>21</sup>

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. These include the remains of ancient Hittite, more fragmentary 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*.<sup>22</sup>

According to the Turkish Statistical Institute, as of 2011 the city of Ankara had a population of 4,338,820 and its metropolitan municipality 4,550,662.<sup>23</sup>

Contents [edit]

- 1 Etymology
- 2 History
- 2.1 Ancient history
- 2.2 Celtic history
- 2.3 Roman history
- 2.4 Turkish history
- 3 Climate
- 4 Landscapes
- 5 Economy
- 6 Main sights
- 6.1 Museums
- 6.2 Attractions
- 6.3 Ankara Castle

BİL 546 - Semantic Web – TOBB ETÜ

25

live.dbpedia.org/page/Ankara

**About: Ankara**

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

**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.
dbpedia-owl:areaCode	▪ 312
dbpedia-owl:areaTotal	▪ 251600000.00000 (xsd:double)
dbpedia-owl:country	▪ dbpedia:Turkey
dbpedia-owl:elevation	▪ 938.00000 (xsd:double)
dbpedia-owl:isPartOf	▪ dbpedia:Ankara_Province
dbpedia-owl:isPartOf	▪ dbpedia:Central_Anatolia_Region
dbpedia-owl:leaderName	▪ dbpedia:Meih_Çökçük

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

## 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') .
}
```

# Linked Data

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

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

# 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

## 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/>