Introduction to Semantic Web

Angelica Lo Duca
IIT-CNR
angelica.loduca@iit.cnr.it

Linked Open Data: a paradigm for the Semantic Web

Course Outline

- Introduction to SW
- Give a structure to data (RDF Data Model)
- Represent data through a vocabulary (RDFS)
- Query Data (SPARQL)
- Linked Data
- Setup of a Linked Data Node
- Data Linking
- Data Quality

Course Material

http://wafi.iit.cnr.it/angelica/teaching/2017_sw_phd/

Exam/Goals

- The exam consists in the creation of a Linked Data node.
- It can be done during the lectures, or after the course.
  - **search for a dataset**
    - from your everyday life, e.g. family tree
    - from your work, e.g. data extracted from sensors
    - from your hobbies, e.g. songs
    - ...

Exam/Goals (cont.)

- **Give a structure to your data**
  - define a data model (RDF)
  - define/use a vocabulary (RDFS)
- **Link your data to external sources**
  - DBpedia,
  - GeoNames
  - other Linked Data nodes
- **Publish your data as Linked Data**

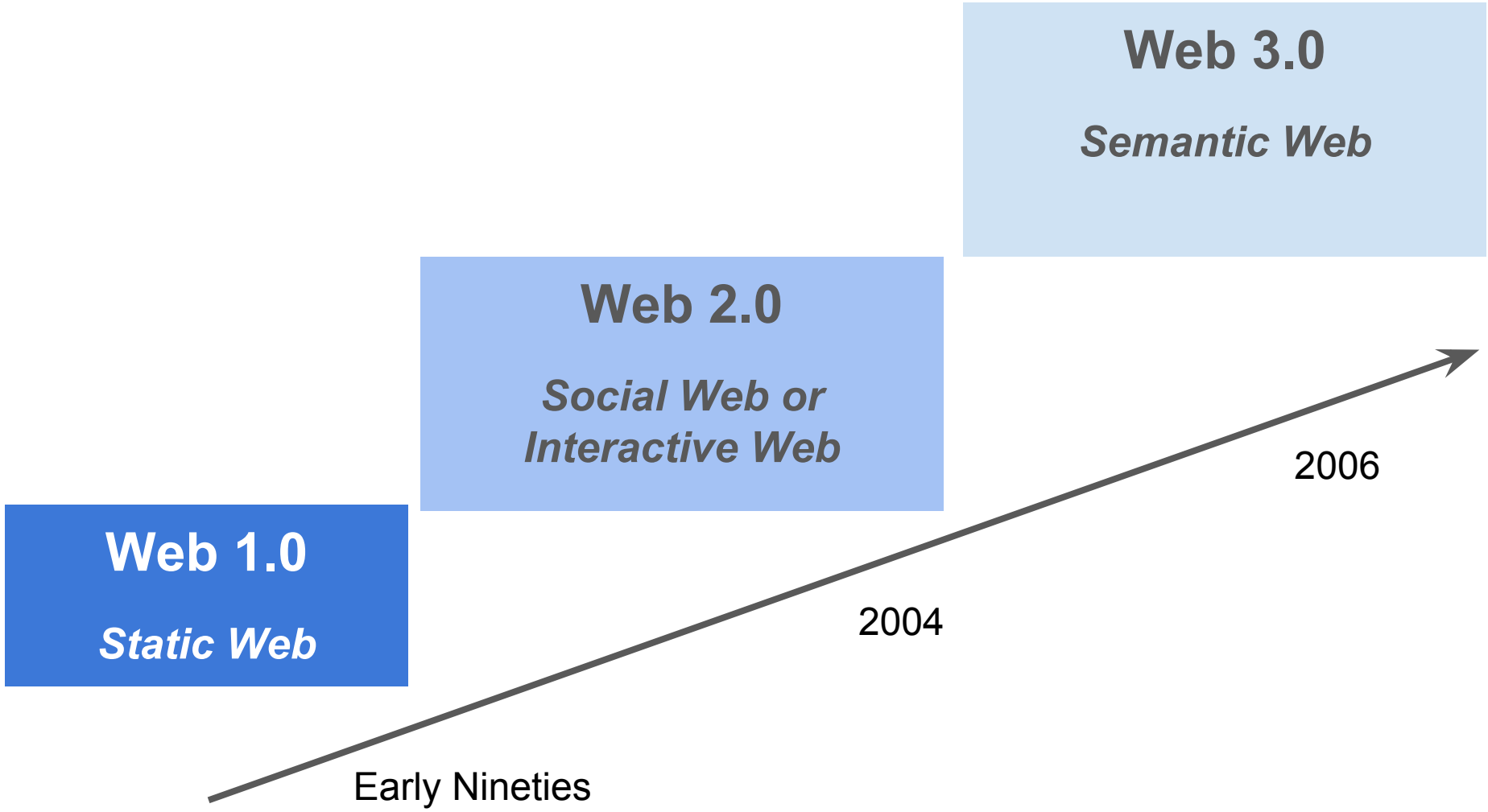
The World Wide Web

Knowledge Base

texts, images, videos,...

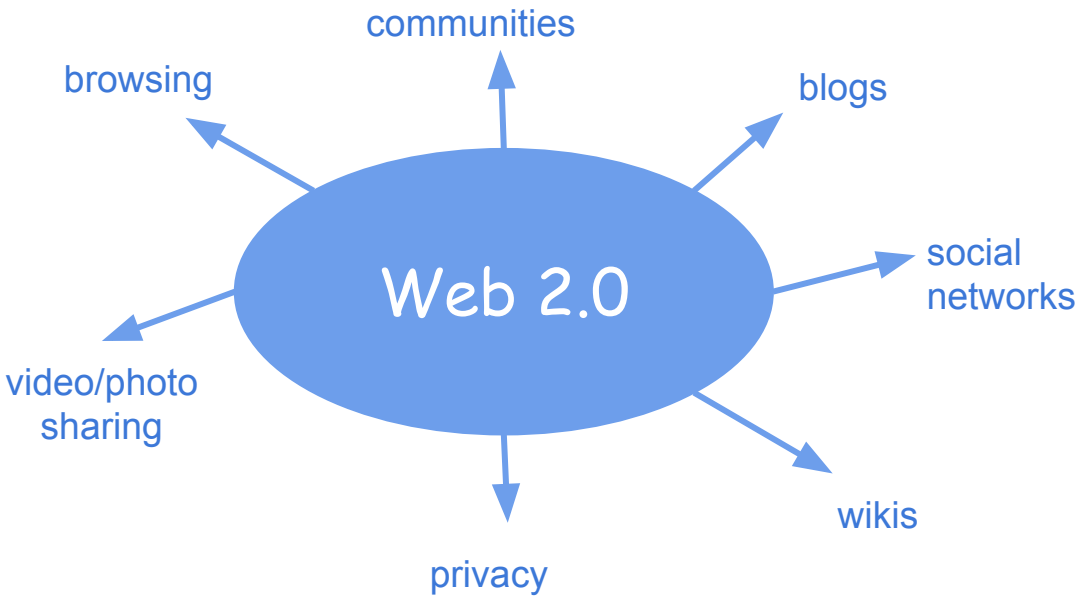
Protocol

to access and generate the Knowledge Base


Web 3.0
Semantic Web
Web 2.0
Social Web or
Interactive Web
Web 1.0
Static Web
2006
2004
Early Nineties

Each web site is a collection of **web pages**

Each web page can refer to one another with global links called **Unifrom Resource Locators** (URLs)



Web 2.0 = Web of documents

Limitations of Web 2.0

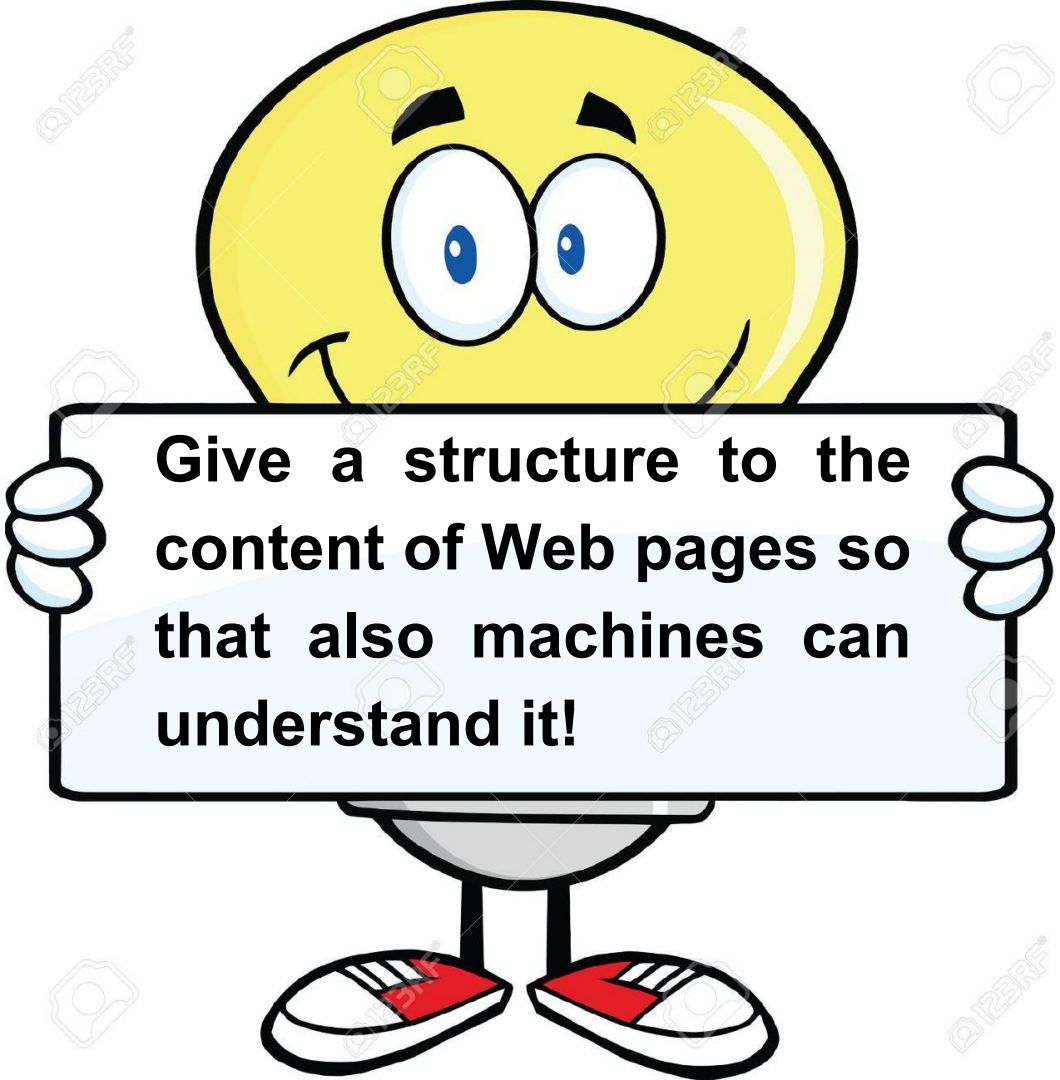
- Too much information with too little structure
- Finding information is not easy
- Data aggregation and reuse
- Data integration from different sources
- Inference of new information

Limitations of Web 2.0 (cont.)

- Web content is heterogeneous
  - in terms of content
  - in terms of structure
  - in terms of character encoding

How to overcome

all these limitations?


Give a structure to the content of Web pages so that also machines can understand it!

“The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

[Tim Berners-Lee et al. 2001.]

Web 2.0 VS Web 3.0

	Web 2.0	**Web 3.0**
granularity	Web of documents	Web of data
target consumers	humans	machines

Web of Data

- Granularity: resource
  - **resource**: everything that has an identity.
  - a web resource is a structure accessible on the web
- Target consumers: intelligent agents (machines)
- Integration & reuse: easier
  - Resources have unique identifiers - Uniform Resource Identifiers (URIs)

The Semantic Web (Web 3.0)

The Semantic Web defines a formal knowledge expressed in a f**ormal language** having:

- a machine-readable notation
- a formal syntax that is strongly coupled with the web architecture
- a formal semantics that provides an access mechanism.

Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. Scientific American Magazine, 2001.

Syntax VS Semantics

- **Syntax** concerned with arrangement of symbols
- **Semantics** concerned with the relation between symbols strings and the world: what things actually *mean*

```
int x = "five";
```

syntax is okay - type identifier = value

semantics is wrong - "five" is not an int

What are the main aims of the SW?

- Automated query-answering
- Automated use of the data (reasoning, planning, acting, etc.)
- ...

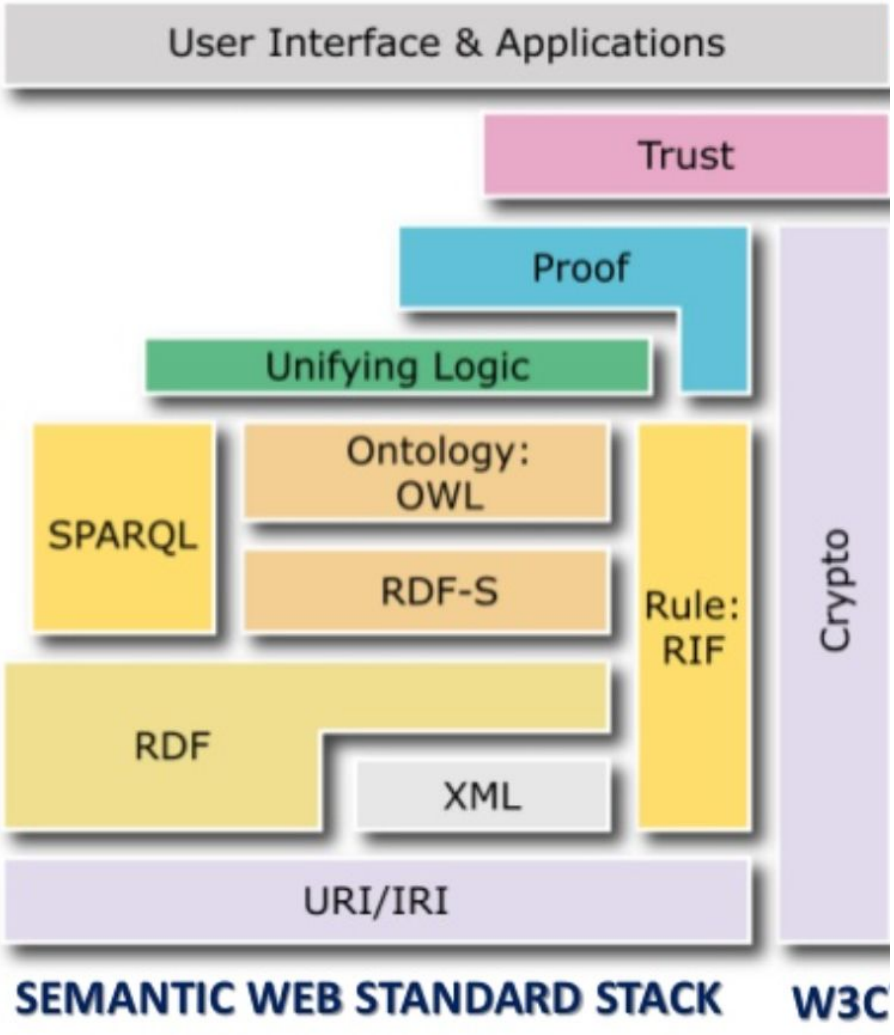
The Semantic Web vision (W3C)

- 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

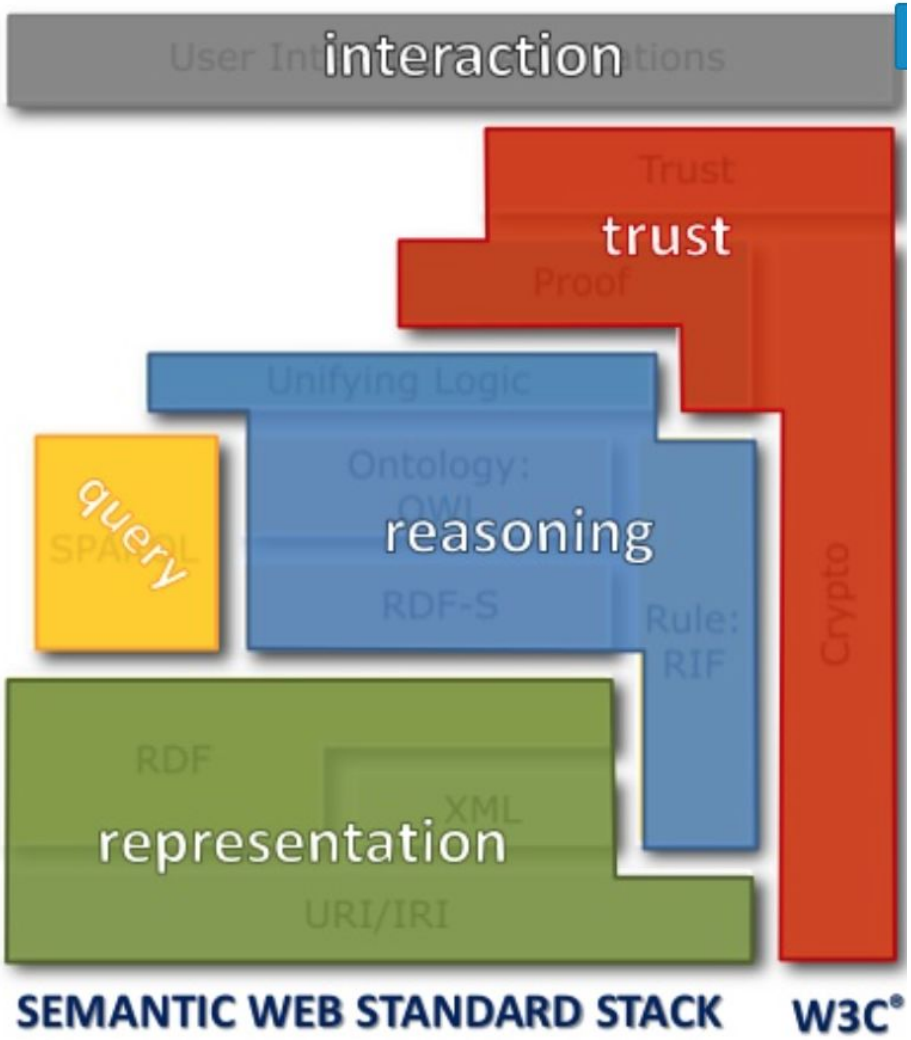
Three strategies

- Introduce concepts of artificial intelligence, to make machines capable of reasoning
- Motivate companies to publish their data freely, using standards defined by Semantic Web
- Promote data reusage.

The SW Layer Cake



The SW Layer Cake



Representation

- **URI/IRI**
  a. Provides means for uniquely identifying semantic web resources
- **Extensible Markup Language (XML)**
  a. markup language for the creation of documents with structured data
  b. language for pure notation
  c. no semantics
- **Resource Description Framework (RDF)**
  a. standard model for data exchange on the web

Reasoning

- **RDF Schema** (RDFS) provides basic vocabulary for RDF.
  a. Using RDFS it is for example possible to create hierarchies of classes and properties.
- **Web Ontology Language** (OWL) extends RDFS by adding more advanced constructs to describe semantics of RDF statements.
  a. It allows stating additional constraints, such as for example cardinality, restrictions of values, or characteristics of properties such as transitivity.
  b. It is based on description logic and so brings reasoning power to the semantic web.

Query

- **SPARQL**
  - it can be used to query any RDF-based data
  - similar to SQL

RDF + SPARQL = Linked Data

Trust and Interaction

- **Trust**
  - should implement all the stuff related with trust and security
- **Interaction**
  - should implement all the applications to enable humans to use Semantic Web applications

Both layers:

- contain technologies not yet standardized or
- contain just ideas that should be implemented

References

Tom Heath and Christian Bizer (2011) *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Dean Allemang and James Hendler. 2011. Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL (2 ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Bob DuCharme. 2011. Learning SPARQL. O'Reilly Media, Inc.

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, Foundations of Semantic Web Technologies. Chapman & Hall/CRC, 2009. ISBN: 9781420090505.