Introduction to the Semantic Web
The Web Today

• **Universal resource identifiers (URIs)** to identify documents.
• The **Hypertext Transfer Protocol (HTTP)** to exchange documents between a client and a server.
• **HTML** for marking up information to be presented to human readers through a browser.
• **Search Engines (Google!)** to discover information.
Using Search Engines – Example I

• Assume that you want to learn about HTTP. What can you do?
• Google it!
Example I (cont’d)
Example I (cont’d)

• So search engines work very well, but are sensitive to vocabulary and have problems associated to precision and recall (e.g., high recall – low precision or low/no recall).
Search Engines – Example II

• Assume that you want to buy the book “Semantic Web Primer”.
• Google it!
Example II (cont’d)
Example II (cont’d)
Search Engines – Example III

• Assume now that you want to buy:
  – The **cheapest** copy of the book “Semantic Web Primer” or
  – The copy that will **reach you earlier**.

• You can still use Google and your browser. But **how many clicks** do you need?

• How about using some sort of shopbot e.g., [http://www.kelkoo.co.uk/](http://www.kelkoo.co.uk/)?
Towards a Better Web!

• The **Semantic Web** vision articulated in a Scientific American article by Tim Berners-Lee, James Hendler and Ora Lassila (May 2001).
  – “The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where agents roaming from page to page readily carry out sophisticated tasks for users.”

• You can find the article on various Web sites e.g., [http://www.dcc.uchile.cl/~cgutierrez/cursos/IC/semantic-web.pdf](http://www.dcc.uchile.cl/~cgutierrez/cursos/IC/semantic-web.pdf)

• Notice the words **meaning** (semantics) and **agents** (a role for all of Artificial Intelligence).
How Can we Achieve the Semantic Web?

• Instead of publishing information to be consumed by humans, publish machine-processable data and metadata using terms/languages that can be understood by machines.

• Build machines (agents) that will search for, query, integrate etc. this data.

• Make sure all agents understand your terms/languages.
The Semantic Web Vision Today

• From the Introduction section of the Semantic Web activity of the W3C
  http://www.w3.org/2001/sw/

  “The Semantic Web is a web of data. There is lots of data we all use every day, and it is not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

  Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

  The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.”
The Semantic Web Vision Today (cont’d)

• Stressing the growing need for **Web data integration**. Lots of important Web applications demand this e.g., e-science and e-government.

• Stressing the need for **Web standards/languages for expressing shared meaning**. This is important if we want agents that are not handcrafted only for particular tasks to be developed.

• See the paper “The Semantic Web Revisited” by Nigel Shadbolt, Wendy Hall and Tim Berners-Lee at http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisited.pdf.
The Semantic Web Layer Cake
Emphasis of this Course

• The Semantic Web topics that we will cover in this course are:
  – Linked data
  – Ontologies (and rules)
Linked Data

• **Linked data**: a set of best practices for publishing and connecting structured data on the Web (from [http://linkeddata.org/](http://linkeddata.org/)).

• **Key technologies**:
  – URIs (a generic means to identify entities or concepts in the world)
  – HTTP (a simple yet universal mechanism for retrieving resources, or descriptions of resources),
  – RDF (a data model for structuring and linking data that describes things in the world).
The Linked Data Cloud

Knowledge Technologies
Manolis Koubarakis
Ontologies

• An **ontology** is a formal, explicit, shared specification of a conceptualization of a domain (Gruber, 1993).

• **Conceptualization**: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.

• The term **ontology** is borrowed from Philosophy, where ontology is a systematic account of existence (what things exist, how they can be differentiated from each other etc.).

• Ontologies are typically expressed in some formal logic-based language (e.g., first-order logic). In the Semantic Web, ontologies are usually expressed in ontology languages such as RDFS and OWL that are based on **individuals (or objects)**, **concepts (or classes)**, **relations** and **axioms**.
Outline of the Course

• The Resource Description Framework (RDF and RDFS)
• SPARQL: A Query language for RDF and RDFS
• Description logics
• The Web Ontology Language (OWL 2)
• Ontology Engineering
• Rule languages for the Semantic Web
• Other topics can be covered in your presentations.
Readings

• The papers on the Semantic Web vision mentioned in the presentation.


• The W3C Semantic Web Activity web site (http://www.w3.org/2001/sw/). Browse! There are many interesting introductory tutorials. You should definitely see the recent tutorial by Ivan Herman http://www.w3.org/2010/Talks/0622-SemTech-IH/Tutorial.pdf which I will use in class to give you an idea of the current state of the art in this area.