The Semantic Web Revisited

Nigel Shadbolt
Tim Berners-Lee
Wendy Hall
Today’s web

- It is designed for human consumption
- Information retrieval is mainly supported by keyword-based search engines
- Some problems with information retrieval:
  - High recall, low precision
  - Low or no recall
  - Results are highly sensitive to vocabulary
  - No integration of multiple sites – results restricted to single pages

⇒ Web content is not *machine-processable* and computers cannot understand and interpret the contents
Semantic Web

• An evolving extension of World Wide Web in which web content can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily.

• Proposed by W3C(World Wide Web Consortium) Chairman T.Berners Lee
For example: Plan a trip to Boston

Current Web
Use search engine to list airlines.
Check each airline for suitable flights with cheapest price, and decide the airline company.
Make a reservation.
Use search engine to list hotels at Boston.
Check each hotel, decide the hotel and make a reservation.
Print out the information about flight, and hotel.

Semantic Web
Ask to list available flights to Boston for any airline with low price. Pick one and it automatically reserve the flight.
Then, it automatically search hotels at Boston which is convenient for the business. Pick one and it automatically reserve the hotel.
All the information is put into your handheld device.
It may also put restaurant phone numbers and other useful information into the handheld device.
Semantic Web – a Web of Metadata

- A web of Metadata – Data about the details of the different web contents, datasources and their relationship in the www
- Data Integration at Web Scale
  - framework for integrating multiple data sources of diverse structures and semantics to draw new conclusions
  - architecture for describing all kinds of things (items, collections, services, processes, etc.)
- Increase the utility of information by connecting it to its definitions and its context
  - effective management and reuse of data at various scales (personal, group, enterprise, community, web)
- Achieved by using three descriptive techniques
  - Resource Description Framework (RDF)
  - Web Ontology Language (OWL),
  - the data-centric, customizable Extensible Markup Language (XML)
Resource Description Framework (RDF)

• A Framework for describing and interchanging the information in the web

• Defines structure (syntax) and semantics of Metadata of the web.

• Builds statement about each resource and their relationship in the form of “Triple”

• Triple (Subject – Predicate – Object expression) based representation language for resources
  • Subject denotes the resource about which statement is made
  • and predicate denotes the aspect of resources and expresses the relationship b/w subject and objects.
  • Object denotes the resource which is the target of predicate

  Eg: “The Sky has the color blue”

  Subject : The Sky
  Predicate : color
  Object : blue
Resource Description Framework (RDF)-RDF Graph

- Graphical representation of RDF triples – *subject node, predicate arc, object node*
- Defines an RDF graph of nodes and labeled directed arcs

$\text{Sky} \rightarrow \text{color} \rightarrow \text{Blue}$
Resource Description Framework (RDF)-Universal Resource Identifiers (URI) & XML

To make these statements *machine-processable* two things are needed:

- a set of *machine-processable* identifiers (for subjects, predicates and objects) without any possibility of confusion between similar looking identifiers

  Uniform Resource Identifiers (URI) allow to uniquely identify the resources such that anyone can link to it, refer to it, or retrieve a representation of it

- a *machine-processable* language for representing these statements and exchanging them between machines

  RDF defines a XML markup language, named RDF/XML, which allows to represent RDF statements.
http://www.example.org/index.html has a creator whose value is John Smith

RDF/XML representation of above graph

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:exterms="http://www.example.org/terms/">

   <rdf:Description rdf:about="http://www.example.org/index.html">
      <dc:creator rdf:resource="http://www.example.org/staffid/5232"/>
   </rdf:Description>

</rdf:RDF>
```
RDF Schema

- RDF’s Vocabulary description language
  - Extends the RDF specification to support the expression of structured vocabularies
- RDF can define the predicate of resources but no mechanism to describe what that predicate and resources are. i.e., the properties and type of these resources.
- RDF Schema provides a higher level of abstraction than RDF.
  - specific classes of resources
  - specific properties,
  - and the relationships between these properties and other resources can be described.
- RDFS allows specific resources to be described as instances of more general classes.
- An RDF & RDFS defines
  - Individuals
  - Kind of things
  - Properties of those things
  - Values of those things
- Similar to object-oriented concept
Triple Stores

- Repository for storing RDF statements
- Necessary for reliable and standardized data access into RDF in the stores
- SPARQL - a protocol and query language designed to fulfill this requirement
RDF Translation

- Need to extract RDF from XML and XHTML documents
- GRDDL (Gleaning Resource Description from Dialects of Languages)
Ontology

- RDFS is recognized as an ontology language
- Ontology is a framework in which it defines a set of concepts for a particular domain and the relationship between these concepts
Web Ontology Language (OWL)

- standardized ontology language that builds upon existing concepts of RDF / RDFS
- adds more vocabulary for describing properties and classes
- W3C provides three Versions of OWL depending the degree of expressivity and complexity
  - OWL Lite, OWL-DL, and OWL Full
- Provides Efficient representation of ontologies that are also amenable to decision procedures
- Checks an ontology to see whether it is logically consistent or to determine whether the particular concept falls within the ontology.
- Uses linking provided by RDF to allow ontology to be distributed across systems
- Allow ontology to refer to terms in other ontologies
OWL Enables Machines to Understand Data!

OWL enables machine-processable semantics!
Structure of the Semantic Web
The Layer Model

• The Layer Model consists of 7 layers.

• The W3C defined/defines a standard for each layer.

• Every layer can access the functionality of the layers below.

• Every layer extends the functionality.
Structure of the Semantic Web
A Possible Architecture
Conclusions

- Today‘s web and its problems
- A vision for a possible semantic web application
- Overview of important standards
  - RDF / RDF Schema
  - OWL