

Introduction to the Semantic Web

Summer School

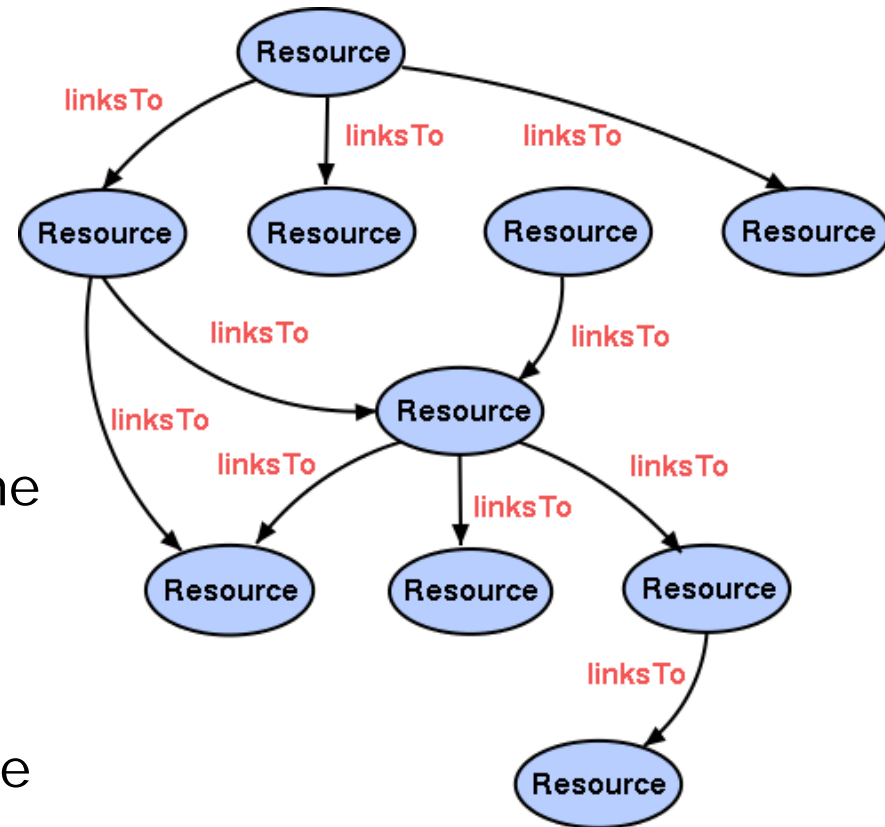
Applications of 3D Shapes: Ontologies, Software Tools and
Industrial Case Studies

July 19-25, 2006, Tallinn, Estonia

Marios Pitikakis (ITI)

The Current Web

- Information and its presentation are mixed up in the form of HTML documents using
 - natural language
 - graphics, multimedia, page layout
- User:
 - exciting world, however, there are no semantics of the resources
- Machine:
 - very little information available - significance of the links only evident from the surrounding text



“Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully.”

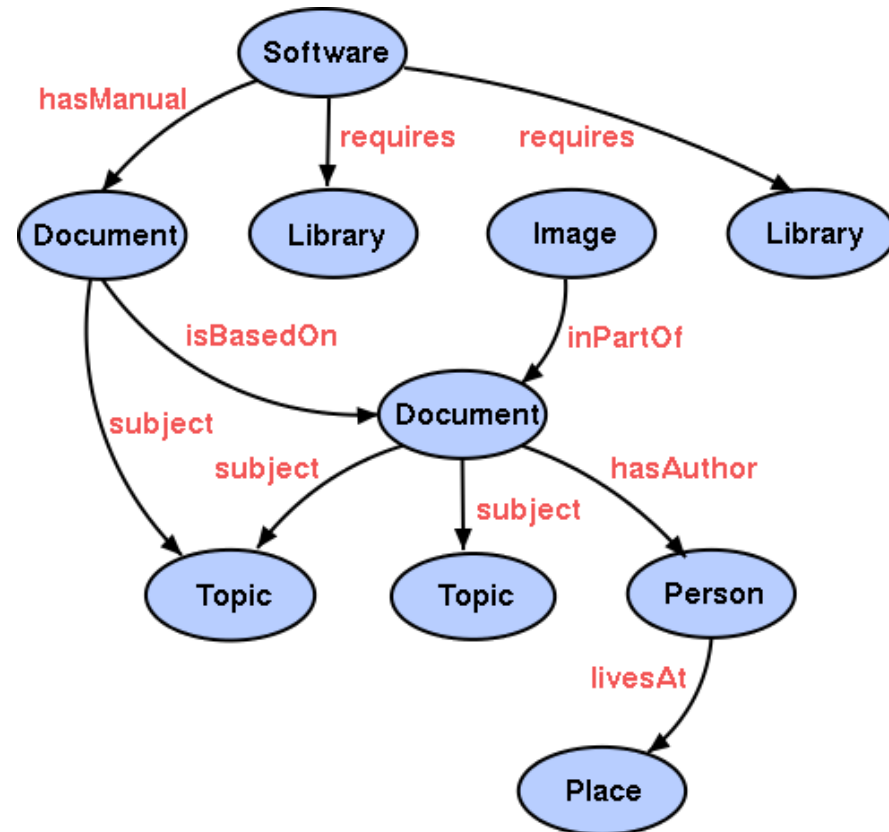
Berners-Lee, T, Hendler, J & Lassila, O 'The semantic web',
Scientific American, May 2001

How can machines understand?

- Content is machine-understandable if it is bound to some formal description of itself (i.e. metadata).

The Semantic Web

- Links:
 - Extensible
 - Relational
- User:
 - Even more exciting world, richer user experience
- Machine:
 - More processable information is available
- Computers and people:
 - Work, learn and exchange knowledge effectively

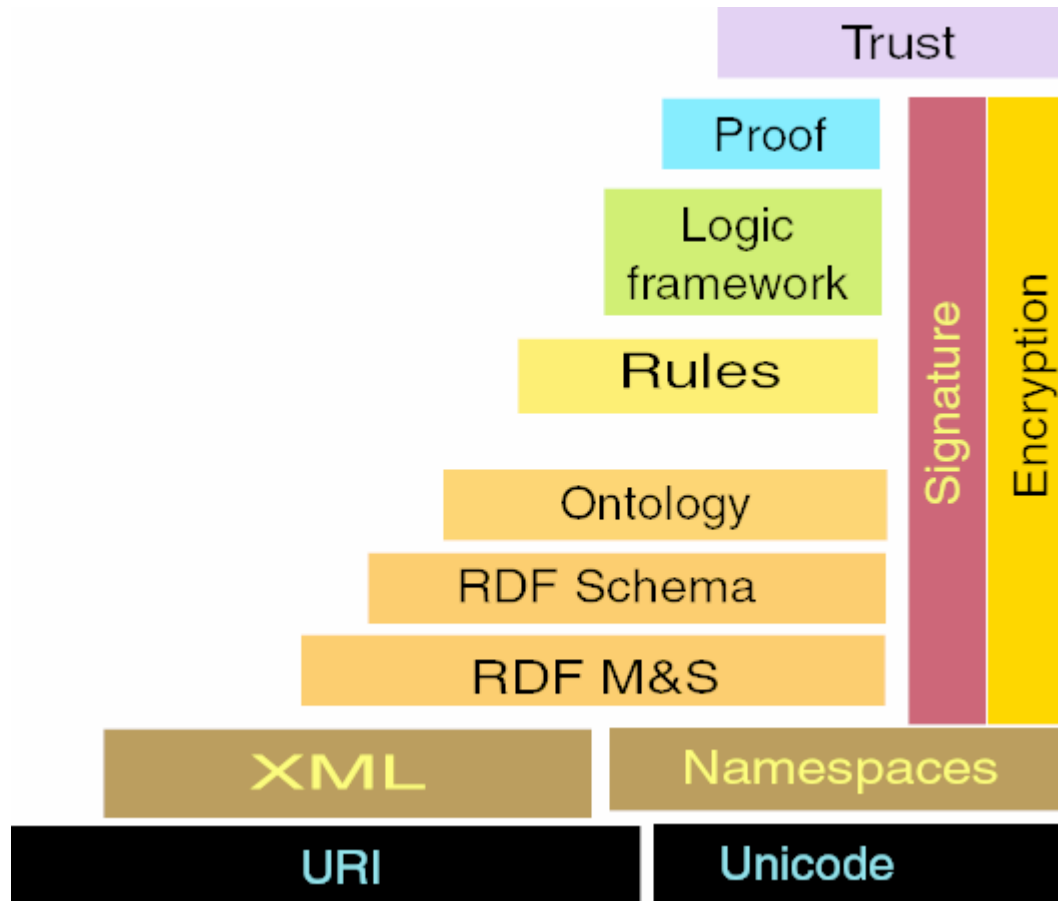


- HTML markup reflect document presentation, but *it cannot adequately represent the semantics & structure of data.*
- But modern Web Applications need more! E.g.
 - Digital Museums & Libraries
 - Electronic commerce
 - On-line Catalogs & Procurement, Comparison Shoppers, Market Places, Virtual Enterprises
 - Scientific applications
 - Advanced Information Management
 - finding, extracting, representing, interpreting, maintaining

- The future of the Web is a universal medium for the exchange of data
- This means applications should have access to data for:
 - interconnection of personal information management
 - personalized services (e.g. news)
 - integrating heterogeneous Web databases
 - better search engines
 - adaptive presentation of information (e.g. for PDAs, phones)

- A resource should provide *information* about itself
 - also called “metadata”
 - metadata stored in XML or other, machine readable form
 - metadata vocabularies should be defined
 - be able to “reason” about (meta)data
- *The “Semantic Web” is a metadata based infrastructure for reasoning on the Web*
- It *extends* the current Web (and does not replace it)

- The “Next Generation Web” aims to provide infrastructure for expressing information in a **precise, human-readable, and machine-interpretable** form
- Enable both **syntactic** and **semantic/structural interoperability** among independently developed Web applications
- Enable Web resources (data & applications) to be **accessible by their meaning** rather than by keywords and syntactic forms
 - Conceptual Navigation & Querying
 - Inference Services (e.g. Picasso is an Artist)



Limitations of XML

- Many ways to say the same thing. Multiple valid structures for the same data
- Not impose a common interpretation of data e.g.
 - heading vs. title
 - price vs. cost
- XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean

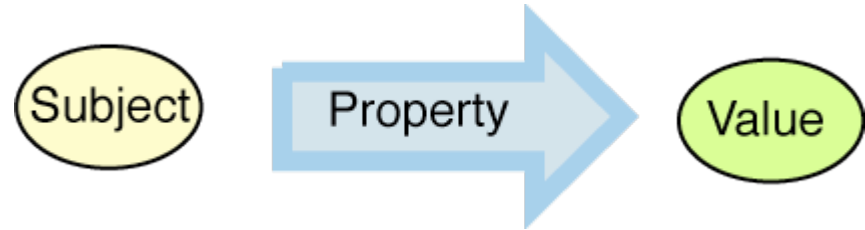
- RDF enables communities to describe their resources in a quite, natural and flexible way
- Data Model: Directed Labeled Graphs
 - **Nodes**: Resources (URIs) or Literals
 - **Edges**: Properties – Attributes or Relationships
 - **Statement**: assertion of the form *resource, property, value*
 - **Description**: set of statements concerning a resource
- XML syntax

- There are three components in an ***RDF triple*** :
 - the subject
 - the predicate (property)
 - the object
- Properties
 - A specific aspect, characteristic, attribute, or relation used to describe a resource
- Statements
 - A specific resource together with a named property plus the value of that property for that resource is an RDF statement.

Relational databases

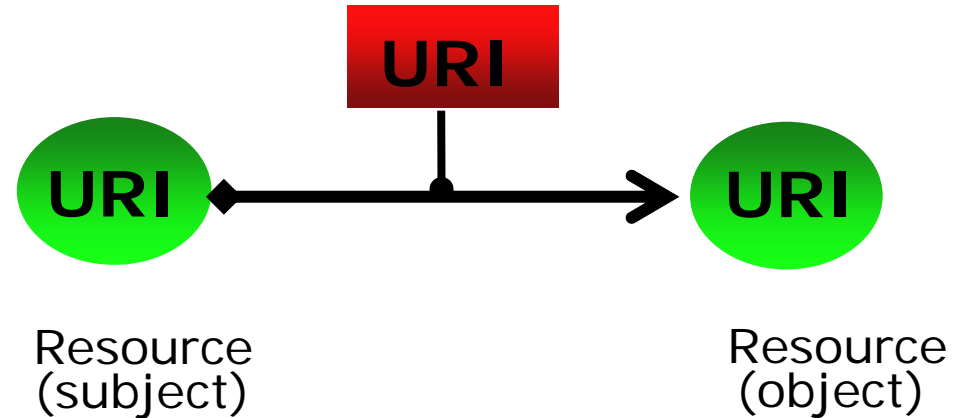
	Property	
Subject	Value	

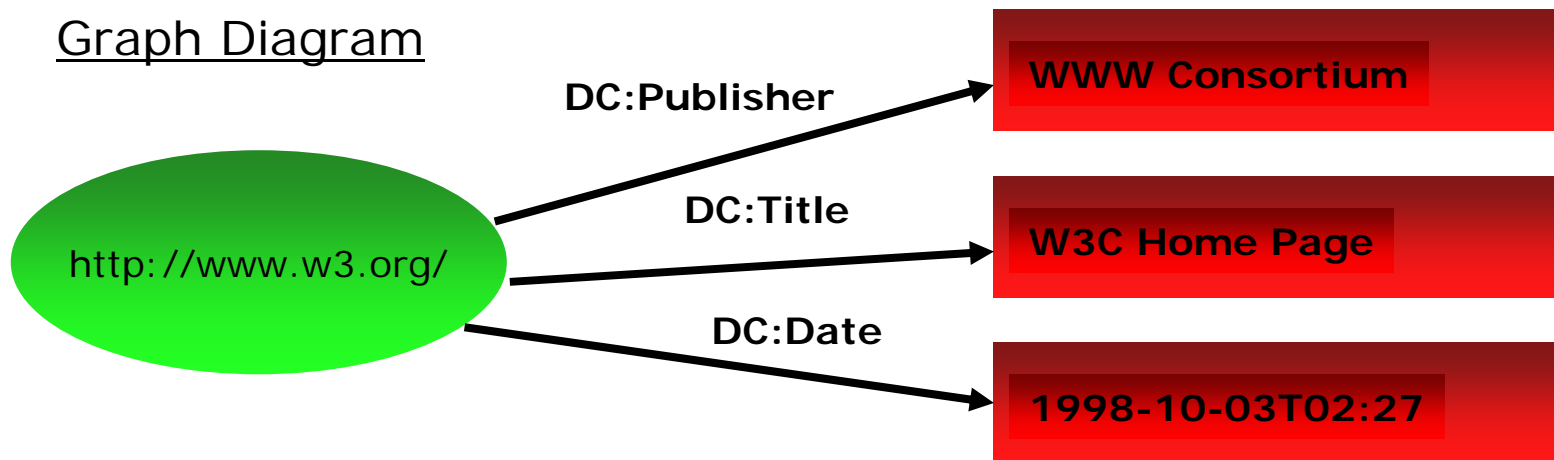
RDF



Statement

Property (Predicate)



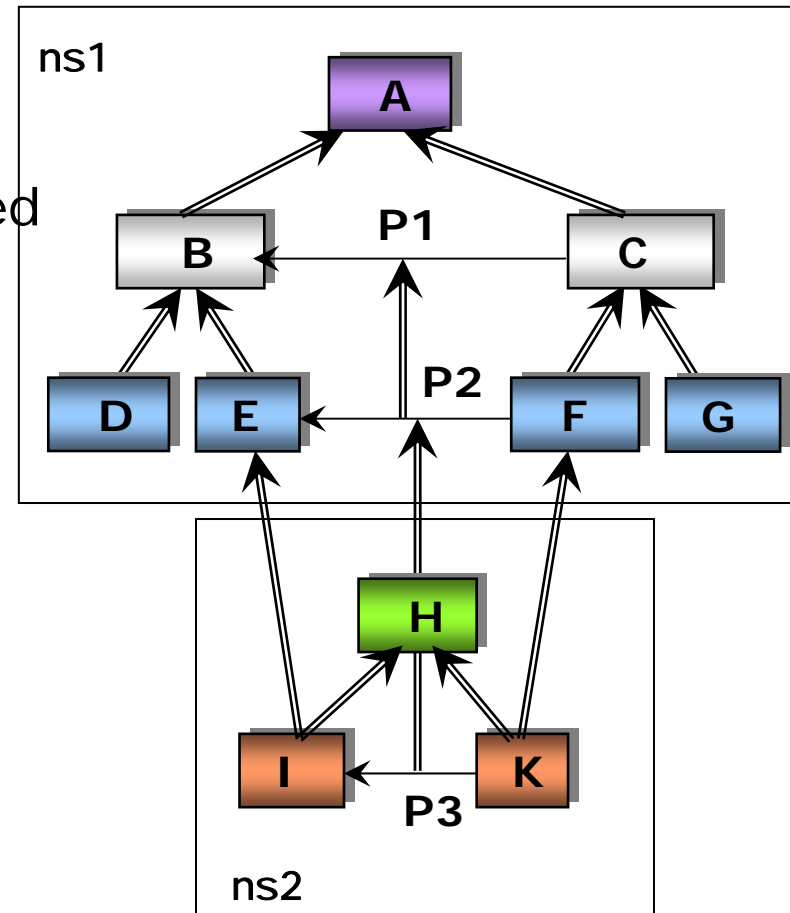
Graph DiagramXML Serialization

```
<rdf:RDF>
  <rdf:Description about="http://www.w3c.org">
    <DC:Publisher>World Wide Web Consortium</DC:Publisher>
    <DC:Title>W3C Home Page</DC:Title>
    <DC:Date>1998-10-03T02:27</DC:Date>
  </rdf:Description>
</rdf:RDF>
```

- XML was designed for documents, not data.
 - Many features (like attributes and entities) are document-oriented, not for expressing data
 - There are many ways to say the same thing in XML
 - Hybrid tree structure: confusing and nonstandard
 - Makes basic operations more complex (e.g. merging)

- RDF was designed for statements, or data
 - Simple structure: triples
 - Merging two documents are simply combining two into one

- *RDF Schema* defines a set of modeling primitives for hierarchically structured vocabularies for machine-processable semantics of information.
 - **Node labels** are defined as **classes**
 - **Edge labels** (predicates) are defined as **properties** of these classes
 - domain and range constraint
 - **Subsumption** of both classes & properties (*subClassOf* and *subPropertyOf*)
 - simple & multiple *is_A*
- RDFS is **expressible in the basic RDF model and syntax**



- The Semantic Web needs the support of *ontologies*:
 - ontologies define the concepts and relationships used to describe and represent an area of knowledge
- We need a *Web Ontology Language* to define:
 - the terminology used in a specific context
 - possible constraints on properties
 - the logical characteristics of properties
 - the equivalence of terms across ontologies
 - etc

- Extends RDF Schema to allow ontologies with more complex relationships (e.g.):
 - limit the properties of classes with respect to number and type
 - means to infer that items with various properties are members of a particular class
- OWL adds value to the Semantic Web through
 - Allowing new data to be inferred
 - Allowing automatic detection of contradictory claims
 - Exploiting existing tools from Description Logic community (e.g. DL reasoners)

- The following are some commercial products that use W3C Semantic Web technologies.
- For more information see <http://www.w3.org/2001/sw/#commercial>

"Oracle Database 10g Release 2 is the world's first mainstream commercial database to provide direct and native support for Semantic Web technologies from W3C.

Oracle has already enabled many customers in the healthcare, life sciences and government industries to use semantic technologies to create exciting new solutions to real-world problems. Taking advantage of the scalability, security and availability of the Oracle technology, these new semantic capabilities are certain to find wide application in commercial enterprises."

- Ken Jacobs, VP Product Strategy, Oracle.

Some features

- Native RDF store
- Converters / loaders for existing RDF data
- RDF Query - An RDF_MATCH function which can be used in SQL to find graph patterns in RDF (similar to SPARQL)

*“The Vodafone Live! service is a 'mobile portal' which provides a unified user experience and access to content and servers from a number of providers. Bringing goods closer to the customer increased sales, and **the use of RDF metadata was a key factor in making this possible.** On the Mobile platform, Semantic Web technologies are driving user experiences which touch and enrich millions of people's lives.”*

- Dan Applequist, Senior Technology Strategist,
Vodafone Group Services Limited

- Search application (e.g. ringtones, games, pictures) using RDF
 - better search : page views per download decreased 50%
 - increased revenue : ringtones up 20% in 2 months
- Bringing goods closer to customers increased sales
- RDF was key factor in making this possible

*"As the leading provider of content creation tools to help people communicate better, adding intelligence to media via metadata was integral to our strategy. We developed Adobe XMP (Extensible Metadata Platform) **based on W3C's RDF**, because it provided a flexible and interoperable framework for fostering the capture, preservation, and interchange of metadata across digital media and workflows. The Adobe Creative Suite provides a design platform that enables creative professional to create information rich assets powered by XMP that can be more effectively repurposed and consumed across multiple media and diverse domains."*

- David Burkett, Director of Product Management,
Adobe Systems

- Adobe's eXtensible Metadata Platform ([XMP](#)) is a labeling technology that allows you to embed metadata into the file itself. With XMP, desktop applications and back-end publishing systems gain a common method for capturing, sharing, and leveraging this metadata valuable opening the door for more efficient job processing, workflow automation, and rights management, among many other possibilities.
- RDF/XML cross product metadata toolkit, supporting the creation, processing, and interchange of document metadata across publishing workflows.
- Provides effective management of digital resources.

- The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework ([RDF](#)), which integrates a variety of applications using XML for syntax and URIs for naming.
- The Semantic Web is about two things. It is about common formats for interchange of data, where on the original Web we only had interchange of documents. Also it is about language for recording how the data relates to real world objects.

Additional information

- W3C World Wide Web Consortium:
<http://www.w3.org/>
- W3C Semantic Web Home Page:
<http://www.w3.org/2001/sw/>