7.10 XML Applications: Semantic Web

- XML-based languages developed for several different application areas; For example
- Semantic Web
  - vision of future Web where applications would process information automatically
  - requires explicit encoding of semantics
- Resource Description Framework, RDF is a general model of metadata for describing (web) resources


Resource Description Framework, RDF

- for describing resources on the web, and for exchanging information about web resources
- designed to be read by computer
- uses URIs to identify web resources (or anything identifiable)
- uses properties with values to describe resources
- RDF also uses the terms subjects, predicates and objects
- concrete syntax RDF/XML based on XML and Namespaces
  - The RDF Schema language allows
    - to describe RDF vocabularies, and especially
    - to provide information about the interpretation of the statements given in an RDF data model.

RDF: Motivation

- Web metadata: providing information about Web resources and systems that use them.
- To do for machine processable information (application data) what the World Wide Web has done for hypertext:
  - to allow data to be processed outside the particular environment in which it was created, in a fashion that can work at Internet scale.
- Interworking among applications: combining data from several applications to arrive at new information.

RDF: Background & Specifications

RDF builds on ideas borrowed from
- Knowledge Representation (KR) in classical AI
  - esp. semantic nets
W3C Recommendations February 2004 for RDF:
1. Primer
2. Concepts and Abstract Syntax
3. Semantics
4. RDF/XML Syntax Specification
5. Vocabulary Description Language 1.0: RDF Schema
6. Test Cases

Replace the 1999/2000 versions of (2) and (5)

Examples of Use
- Describing properties for shopping items, like price and availability
- Describing time schedules for web events
- Describing information about web pages, like date of creation and modification, title, and author
- Describing content and rating for web pictures
- Describing content for search engines
- Describing electronic libraries
RDF: subjects, predicates and objects

• RDF terminology for talking about parts of statements:
  • The part that identifies the thing the statement is about is called the subject
  • The part identifying the property or characteristic of the subject that the statement specifies is called the predicate
  • The part that identifies the value of that property is called the object.

For example, the English statement
http://www.example.org/index.html has a creator whose value is John Smith.

RDF: subjects, predicates and objects

• the subject is the URL
  http://www.example.org/index.html
• the predicate is the word "creator"
• the object is the phrase "John Smith"
• English is good for communicating between (English-speaking) humans,
• RDF is about making machine-processable statements.
• For that, two things are needed:
  1. machine-processable identifiers for identifying subjects, predicates, and objects without possibility of confusion.
  2. a machine-processable language for representing these statements and exchanging them between machines.

RDF: Data Model

• An RDF expression is a collection of triples
  (subject, predicate, object)
• A set of such triples forms an RDF graph.
  Illustrated by a graphic diagram:
  • The nodes of an RDF graph are its subjects and objects
  • each triple is represented as a node-arc-node link
  • The arc is directed form subject to the object

RDF: Graph Data Model

• Groups of statements are represented by groups of nodes and arcs
• The meaning of an RDF graph is the conjunction (logical AND) of the statements corresponding to the triples it contains.
• So, to reflect the additional English statements
  http://www.example.org/index.html has a creation-date whose value is August 16, 1999
  http://www.example.org/index.html has a language whose value is English
the following RDF graph could be used (using suitable URIs to name the properties "creation-date" and "language"):

Several Statements About the Same Resource

• A relationship, indicated by the predicate, holds between the things denoted by subject and object of the triple.
• A Simple RDF Statement:

Uses Dublin Core metadata properties
RDF: Graph Data Model

• Objects in RDF statements may be either
  – URIs, or
  – constant values (called literals) represented by
    character strings, in order to represent certain kinds of
    property values.
• In the case of the predicate
  http://purl.org/dc/elements/1.1/language the
  literal is an international standard two-letter code for
  English.
• Literals may not be used as subjects or predicates in RDF
  statements.

RDF Example

Few lines from a list of CD records:

<table>
<thead>
<tr>
<th>Title</th>
<th>Artist</th>
<th>Country</th>
<th>Company</th>
<th>Price</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empire Burlesque</td>
<td>Bob Dylan</td>
<td>USA</td>
<td>Columbia</td>
<td>10.90</td>
<td>1985</td>
</tr>
<tr>
<td>Hide your heart</td>
<td>Bonnie Tyler</td>
<td>UK</td>
<td>CBS Records</td>
<td>9.90</td>
<td>1988</td>
</tr>
</tbody>
</table>

• An RDF/XML description of these CD resources:

```xml
<?xml version="1.0"?><rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cd="http://www.recshop.fake/cd#">
  <rdf:Description
    rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
    <cd:artist>Bob Dylan</cd:artist>
    <cd:country>USA</cd:country>
    <cd:company>Columbia</cd:company>
    <cd:price>10.90</cd:price>
    <cd:year>1985</cd:year>
  </rdf:Description>
  <rdf:Description
    rdf:about="http://www.recshop.fake/cd/Hide your heart">
    <cd:artist>Bonnie Tyler</cd:artist>
    <cd:country>UK</cd:country>
    <cd:company>CBS Records</cd:company>
    <cd:price>9.90</cd:price>
    <cd:year>1988</cd:year>
  </rdf:Description>
</rdf:RDF>
```

RDF Model

• From the W3C RDF Validation Service

```xml
<!-- description of a resource identified by
the rdf:about attribute -->
<!-- The cd:artist, cd:country, etc. element
describes a property of the resource -->
<rdf:Description
  rdf:about="http://www.recshop.fake/cd/Hide your heart">
  <cd:artist>Bonnie Tyler</cd:artist>
  <cd:country>UK</cd:country>
  <cd:company>CBS Records</cd:company>
  <cd:price>9.90</cd:price>
  <cd:year>1988</cd:year>
</rdf:Description>
```
RDF: Resources, Properties, Values

- "Resources have Properties with Values".

- The example above says:
  - The Resource: http://www.recshop.fake/cd/Hide your heart" has a Property called "artist" with value "Bonnie Tyler"
  - The Resource: http://www.recshop.fake/cd/Hide your heart" has a Property called "price" with the value "9.90"
  - Natural language: "The CD Hide your heart costs $9.90"

The Description Element

- `<rdf:Description>`
  - The Description element describes a resource.
  - The about attribute identifies the resource.
  - Property elements are used to describe the resource e.g. `<cd:artist>Bob Dylan</cd:artist>`
  - The property elements (artist, country, company, price, and year) are defined in the namespace `xml:cd`.
    - namespace is outside RDF
  - RDF defines only the framework.
  - Properties must be defined by someone else (company, organisation, person, etc).

Properties as Attributes

- Properties can also be expressed as attributes (instead of elements):

Resources as Properties

  - `rdf:resource` indicates a value that is a resource (and not a literal)

RDF Schema

Application-specific classes and their properties can be defined using RDF Schema (RDFS)

- RDF Schema provides the framework to describe application-specific classes and properties
  - C.f. describing application-specific elements and attributes with DTDs or XML Schemas
- Classes in RDF Schema are much like classes in object oriented languages.
- This allows resources to be defined as instances of classes, and subclasses of classes.

RDFS Example

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.animals.fake/animals">
  <rdf:Description rdf:ID="Animal">
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  </rdf:Description>
</rdf:RDF>
```
RDFS Example

```
<rdf:Description rdf:ID="Horse">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:subClassOf rdf:resource="#Animal"/>
</rdf:Description>

<rdf:Description rdf:ID="Dog">
  <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:subClassOf rdf:resource="#Animal"/>
</rdf:Description>
</rdf:RDF>
```

Applications of RDF

- RDF has some real-world applications. Here are some:
  - Dublin Core Metadata
    - widely used elements to document Internet resources, e.g., title, creator, subject, description, date, ...
    - described using RDF Schema
  - Adobe’s Extensible Metadata Platform (XPM)
    - to embed RDF/XML metadata in (PDF) documents
  - RSS 1.0: RDF Site Summary
    - to describe newsfeeds
    - perhaps the most widely deployed RDF app. on Web
  - CC/PP (Composite Capabilities/Preference Profiles)
    - framework for client devices to express their capabilities and preferences to receive content

Beyond RDF

- Standardization for Semantic Web has continued by extended languages based on RDF
- -> DAML+OIL -> OWL (Web Ontology Language)
- to represent and share machine-processable ontologies
  - vocabularies with semantics and relationships of their terms
  - OWL more powerful than RDF, for example, to describe relationships of classes (e.g., equal, disjoint) and cardinality constraints ("exactly one")

XML Applications: Semantic Web: OWL

- OWL is a stronger language with greater machine interpretability than RDF.
- OWL comes with a larger vocabulary and stronger syntax than RDF.
- OWL extends a vocabulary description, allowing to express claims, e.g.
  - Nothing can be both a Document and a Person
  - grandParent and grandChild are inverses
  - homepage, nasdaqCode and mbox are uniquely identifying properties
  - A W3CTeamPerson is a Person whose workplaceHomepage is http://www.w3.org/
  - ...as well as the formally specified interactions amongst all these.

Semantic Web Applications

- Large number of applications emerge (?)
  - first applications were RDF only
  - newer ones begin to use ontologies
    - huge number of ontologies exist already, with proprietary formats
    - converting them to RDF/OWL will be a major task
      - (but there are converters)
  - See, for example, on WSIndex
    - portal on “Web Services and Semantic Web Resources”
Discussion / Speculation

- Does Semantic Web "start to fly"?
- Standardized representations of metadata are surely useful
- Automatic processing requires applications to "understand" the semantics of classes and properties
- The volume of Web may make a difference, if commonly understood vocabularies are used sufficiently often
- Compare: Hypertext was developed for decades in laboratories and stand-alone systems
- The distributed nature and volume of Web "made it to fly"
- Motivation of information providers to provide metadata on top of their content?