The Semantic Web:
The Roles of XML and RDF

Presenter: Litao Deng
CSCI 586 Database Systems Interoperability
Semantic Web is an effort to enhance current web so that computers can process the information presented on WWW, interpret and connect it, to help humans to find required knowledge.

Semantic Web is intended to form a huge distributed knowledge based system (share data instead of documents).

Semantic Web provide a common framework that allows data to be shared and reused across application, enterprise, and community boundaries.
Knowledge Sharing is a key factor in Semantic Web.

Ontology can formally represent knowledge as a set of concepts within a domain, and the relationships among those concepts.
Overview for this Paper

We explain the role of ontologies in the architecture of the Semantic Web.

We summarize key elements of XML and RDF.

We show using XML as a tool for semantic interoperability will be ineffective in the long run.

We give a general method for encoding ontology representation languages into RDF/RDF schema.
Ontology

What is ontology?

An ontology is a **specification** of a **conceptualization**, is a **description** of the **concepts** and **relationships** in a specific domain.

Ontologies provide a **common understanding of topics** that can be communicated between people and application systems.
**OIL** (Ontology Interchange Language) is the language to represent or authoring ontology, consists of slot definitions (**slot-def**) and class definitions (**class-def**).

```plaintext
class-def defined herbivore
   subclass-of animal, NOT carnivores
   slot-constraint eats
      value-type plant
   OR (slot-constraint is-part-of has-value plant)
```

% herbivores are animals, but not carnivores
% that eat only plants or part of plants

One example ontology to use **OIL** to define herbivore.
What We Need?

To achieve **Semantic Web**, we need to use **ontologies** to enable web-based **knowledge** processing, sharing and reuse between applications. (This is called **Semantic Interoperability**).

So we need to use some toolkits to apply ontology representation languages (knowledge) like **OIL** to the task of interoperability. (Encode ontologies (language) into an interoperability form)

**XML** and **RDF** are two candidates.
**XML**: a well formed XML document creates a balanced tree of nested sets of open and close tags, each of which can include several attributed-value pairs.

**DTD & XML schema**: defines a *grammar* to specify allowable combinations and nesting of tag names, DTD specifies *only syntactic* conventions; any intended semantics are outside the realm of the XML specification.
XML

Serialization syntax for other markup language.
Semantic markup of web pages.
Uniform data-exchange format.

One example ontology to use XML serialization to define branch.

```xml
<class-def>
  <class name="branch"/>
  <slot-constraint>
    <slot name="is-part-of"/>
    <has-value>
      <class name="tree"/>
    </has-value>
  </slot-constraint>
</class-def>
```
RDF: designed to standardize the definition and use of metadata – descriptions of web-based resources. RDF is an **object-attribute-value** (object-predicate-value) triple, commonly written as $A(O, V)$.

**In RDF:** Object and value can be **interchanged**. RDF **statement** can be the object or value of a triple.
RDF Schema

**RDF schema** lets developers define a particular vocabulary (such as `authorOf`) for RDF data and specify the kinds of object to which these attributes can be applied. **A basic type system.**

```xml
<rdf:RDF xml:lang="en" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdf:Description ID="MotorVehicle">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  </rdf:Description>
  <rdf:Description ID="Truck">
    <rdf:type resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#MotorVehicle"/>
  </rdf:Description>
</rdf:RDF>
```
Knowledge Representation

Recall that we want to enable web-based knowledge processing, sharing and reuse. Summarize the requirements for data (knowledge) exchange.

*Universal expressive power.*

*Syntactic interoperability.*

*Semantic interoperability.*
Using XML

*Good universal expressive power:* anything for which a grammar can be defined can be encoded in XML.

*Good syntactic interoperability:* an XML parser can parse any XML data.

*Poor semantic interoperability:* XML just describes *grammars*, and there is no way to recognize a *semantic unit* from a particular domain, because XML aims at document structure and imposes no common interpretation of the data contained in the document.
Using XML example

**Fixed and flexible communication:** One domain model, multiple DTD. It is difficult to reengineer the domain model from the DTDs. The data can not (hardly) be exchanged.
Using RDF

Good universal expressive power: RDF’s nested object-attribute-value structure satisfies our universal expressive power requirement.

Good syntactic interoperability: application-independent RDF parsers are also available to fulfills our syntactic interoperability.
Using RDF

Good semantic interoperability:
The object-attribute-value structure provides natural semantic units.

A domain model (ontology or something define objects and relationship) can be represented naturally in RDF.

Techniques to find mappings between RDF descriptions are directly applicable.

RDF model can still be usable even if current XML syntax changes or disappears.
How to enrich RDF

Why?
Enrich RDF to represent more sophisticated and expressive knowledge (data).

Using the primitives from ontology language L to describe a particular domain:
1. Describe language L’s modeling primitives using RDF schema.
2. Describes a specific ontology in L using the resulting RDF schema document.
3. Use the RDF schema documents to describe instances of the specific L ontology modeled in step2.
Merging an ontology language with RDF schema

Ontology language and RDF schema *can not* be 100-percent compatible.

1. Use the RDF schema vocabulary wherever possible.
2. Every OIL expression would be an object.
3. …
Conclusion

Semantic interoperability will be a *sine qua non* for the Semantic Web.

We must exploit the current *RDF* proposals, rather than XML labeling.

The challenge is to expand this generic method for creating web-enabled, special purpose *knowledge representation languages*. 
Status of the Semantic Web (Layer Cake)
Backup Slides

Current Ontology Language for the Semantic Web

\[ \text{RDF/RDFS} \rightarrow \text{DAML/OIL} \rightarrow \text{DAML + OIL} \rightarrow \text{OWL (OWL Lite, OWL DL, OWL Full)} \]