CS 2740 Knowledge representation
Lecture 20

Semantic web

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

Semantic web

The **Semantic Web** provides a common framework that allows data and knowledge to be shared and reused across application, enterprise, and community boundaries.

The development of the Sematic web is a collaborative effort led by **W3C** with participation from a large number of researchers and industrial partners.

It defines standards for exchanging knowledge and for sharing conceptualizations.

Standards:

- **RDF** - the Resource Description Framework, representation of information/data for the purpose of sharing
  (**XML** was used to share documents only)
- **OWL** – a language for sharing vocabularies, sets of terms supporting web searches and other applications (within RDF)

Semantic web

The Semantic Web is really about:

- **common formats for interchange of data and knowledge**, whereas on the original Web we only had to interchange of documents (XML are for exchanging documents)
- **language for recording how the data relates to real world objects**. That lets a person, or a machine, to start off in one database, and then move through set of databases which are connected not by wires but by being about the same thing.

Semantic web

Benefits:

- **knowledge integration,**
- **knowledge storage,**
- **knowledge inference,**
- **and knowledge searching.**

Semantic web: knowledge integration

Benefit of large amounts of information and knowledge on the web stands and falls on the data/knowledge integration

Technical challenges:

- ***Location*:** where the data/knowledge resides? The *location* of a Semantic Web resource is defined by the **Uniform Resource Identifier (URI).** A URI is simply a formatted string that identifies - via name, location, or any other characteristic - a resource. This allows us to label a Semantic Web source with a *findable, unique* location.
- ***Query Protocol*:** We have to be able to interact with the data/resources. We need an exchange language. The protocol for the Semantic Web uses standards such as ***http*** to form a flexible, easily understood, request/response exchange.
- ***Format*:** The data must be in a comprehensive and translatable format. The Semantic Web uses a standard format - the **OWL Web Ontology Language.** It is based on the **Resource Description Framework (RDF)** standard and **XML.**

Technical challenges are resolved by standards

Semantic web: knowledge integration

Other Challenges

- ***Timely, Authoritative*:** The data must be trusted and be up-to-date. Having multiple answers to the same question is rarely fun. Having outdated information or information that requires synchronization further complicates the challenge. The Semantic Web allows you to deal directly with the actual source of record. You need not maintain complex polling and synchronization unless it is absolutely necessary due to performance or other requirements.

The key challenge:

- ***Purpose*:** We have to align the data with our purpose. This may require translation and modifications. It needs to fit your world view be it medical, financial view. This is about getting right the semantic. For example, we can tie *person* in one data source with *individual* from another data source - they represent the same meaning or a related meaning.
- Semantic web standards do enable easier and more efficient data sharing and integration but really reach their full potential by the ability to align purpose across different data sources.

Semantic web:knowledge integration

Three steps of integration:

- ***Aggregation*:**
  - Combines the Semantic Web data sources into one unified, virtual data source.
- ***Mapping*/Binding:**
  - Associates similar references with each other and builds upon data in existing references. For example synonyms are identified.
- ***Rules*:**
  - Enables more sophisticated alignment and enrichment such as conditional logic that adds information based on the condition of other data

Semantic web: OWL

The Semantic Web is always in the same language:

- **The OWL Web Ontology Language (http://www.w3.org/TR/owl-ref/)**

The Web Ontology Language OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of RDF (the Resource Description Framework)

OWL contains all the reference information (self contained or *even* remote) to **define any term contained within** - it maintains its own definition of each and every term (it is self-referential).

Ontology

If more than one person is building a knowledge base, they must be able to share the conceptualization.

- A conceptualization is a mapping from the problem domain into the representation.
- A conceptualization specifies:
  - What types of objects are being modeled
  - The vocabulary for specifying objects, relations and properties
  - The meaning or intention of the relations or properties
- An ontology is a specification of a conceptualization.

Semantic web: OWL

```
- <owl:Ontology rdf:about="">
  <rdfs:comment>This is a weather forecast ontology.</rdfs:comment>
  <rdfs:label>Weather Site Ontology</rdfs:label>
  </owl:Ontology> <!-- Weather Observation Class   -->
- <owl:Class rdf:ID="WeatherObservation">
  <rdfs:label>Weather Observation</rdfs:label>
- <rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasLocation" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
- <rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasTime" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
…
```

Semantic web: OWL

```
…
<rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasTemperature" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
- <rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasHumidity" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
- <rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasWindSpeed" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```


Semantic web: OWL

```
<!-- Location Class  -->
- <owl:Class rdf:ID="Location">
  <rdfs:label>Location: City, State</rdfs:label>
- <rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasState" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
- <rdfs:subClassOf>
- <owl:Restriction>
  <owl:onProperty rdf:resource="#hasCity" />
  <owl:cardinality>1</owl:cardinality>
  </owl:Restriction>
  </rdfs:subClassOf>
  </owl:Class>
- <owl:DatatypeProperty rdf:ID="hasState">
  <rdfs:label>The State that this location is in. Abbreviated.</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string" />
  </owl:DatatypeProperty>
```

Semantic web: OWL

```
<!-- Precipitation Class   -->
- <owl:Class rdf:ID="Precipitation">
  <rdfs:label>Precipitation Condition</rdfs:label>
- <owl:oneOf rdf:parseType="Collection">
  <Precipitation rdf:about="#Snow" />
  <Precipitation rdf:about="#Rain" />
  <Precipitation rdf:about="#Thunderstorm" />
  <Precipitation rdf:about="#None" />
  </owl:oneOf>
  </owl:Class>
```


Semantic web: OWL

```
<!-- Properties   -->
- <owl:ObjectProperty rdf:ID="hasLocation">
  <rdfs:label>Location of observation.</rdfs:label>
  <rdfs:range rdf:resource="#Location" />
  </owl:ObjectProperty>
- <owl:DatatypeProperty rdf:ID="hasTime">
  <rdfs:label>Date and time of observation.</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string" />
  </owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="hasTemperature">
  <rdfs:label>Temperature, farenheit</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#float" />
  </owl:DatatypeProperty>
- <owl:DatatypeProperty rdf:ID="hasHumidity">
  <rdfs:label>Relative humidity, percent.</rdfs:label>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#float" />
  </owl:DatatypeProperty>
```