OWL and Protégé

Hoang Huu Hanh
IFS – TU Wien
Content

- Finding issues
  - Roles of OWL and Protégé 2000
  - Ontology-based search and selection
  - A Formal Approach for Querying the Semantic Web
- Future works
  - Description Logics
OWL - What’s OWL?

- OWL is a set of XML elements and attributes, with standardized meaning, that are used to define terms and their relationships.
- OWL extends RDF Schema:

  - Class
  - equivalentProperty
  - sameIndividualAs
  - ...
  - subClassOf
  - resource
  - ID
  - ...

  OWL elements and attributes (i.e., OWL Vocabulary)

  RDF Schema
OWL (cont’d)

- OWL provides an agreed-upon vocabulary for expressing semantics

Samples of OWL Vocabulary:

subClassOf: this OWL element is used to assert that one class of items is a subset of another class of items.
Example: SLR is a subClassOf Camera.

equivalentProperty: this OWL element is used to assert that one property is equivalent to another.
Example: aperture is an equivalentProperty to f-stop.

sameIndividualAs: this OWL element is used to assert that one instance is the same as another instance.
Example: Vienna University of Technology is the sameIndividualAs TU-Wien.
Symmetric: if P(x,y) then P(y,x)
InverseOf: if P1(x,y) then P2(y,x)
Transitive: if P(x,y) and P(y,z) then P(x,z)
Functional: if P(x,y) and P(x,z) then y=z
InverseFunctional: if P(x,y) and P(z,y) then x=z
allValuesFrom: P(x,y) has y=allValuesFrom(C)
someValuesFrom: P(x,y) has y=someValuesFrom(C)
hasValue: P(x,y) and y=hasValue(I)
cardinality: cardinality(P) = n
minCardinality: minCardinality(P) = n
maxCardinality: maxCardinality(P) = n
equivalentProperty: P1 = P2
intersectionOf: C = intersectionOf(C1, C2, …)
unionOf: C = unionOf(C1, C2, …)
complementOf: C = complementOf(C1)
oneOf: C = oneOf(I1, I2, …)
equivalentClass: C1 = C2
disjointWith: C1 != C2
sameIndividualAs: I1 = I2
differentFrom: I1 != I2
AllDifferent: I1 != I2, I1 != I3, I2 != I3, …

Thing: C1, C2, …, I1, I2, …, P1, P2, …

Summary of OWL Capabilities

Institute for Software Technology and Interactive Systems
OWL (cont’d)

- OWL enables machines to understand data

**OWL enables machine-processable semantics!**
OWL and Interoperability

- OWL plays an important role in data interoperability
- Two necessary conditions for interoperability:
  - Adopt a common syntax: this enables applications to *parse* the data. XML provides a common syntax, and thus is a critical first step.
  - Adopt a means for understanding the semantics: this enables applications to *use* the data. OWL provides a standard way of expressing the semantics.
Example: consider following XML fragment

```
<SLR>
  ...
</SLR>
```

- SLR = *Single Length Reflex (camera)* or *Satellite Laser Ranging* or others…?

What is an SLR?
Meaning (semantics) applied on a per-application basis

**Semantics**: A SLR is a type of Camera. (i.e., SLR = Single Lens Reflex)

**Actions**: These actions must be performed on the SLR data:
- check the aperture.
- check the (lens) size.
- check the cost.
Meaning (semantics) applied on a per-application basis

XML

<table>
<thead>
<tr>
<th>app#1</th>
<th>app#2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semantics</strong>: Code to interpret the data</td>
<td><strong>Semantics</strong>: Code to interpret the data</td>
</tr>
<tr>
<td><strong>Action</strong>: Code to process the data</td>
<td><strong>Action</strong>: Code to process the data</td>
</tr>
</tbody>
</table>
Problems encounter:

- **Duplicate effort**
  - Each application must express the semantics

- **Variability of interpretation**
  - Each application can take its own interpretation
  - Example: Mars probe distance - one application interpreted the data in inches, another application interpreted the data in centimetres.

- **No ad-hoc discovery and exploitation**
  - Applications have the semantics pre-wired. Thus, when new data is encountered an application may not be able to effectively process it. This makes for brittle applications.
Better approach:
- Extract semantic definitions from applications
- Express semantic definitions in a standard vocabulary
Why not hardcode the info that is in an OWL document directly into applications?

Why design your application like this:

And not like this:
The Answer Is…

- There are many reasons for keeping the OWL semantic info separate from the application:
  - **Leverage XML**: an OWL document is an XML document. Thus, XML tools can be used to create, edit, and validate OWL documents.
  - **Extensible**: the OWL document can grow independent of the application.
  - **Reusable**: the OWL document can be used by multiple applications.
  - **Avoid Misinterpretation**: the info in an OWL document is expressed using a well-defined, standardized vocabulary. This helps avoid misinterpretation.
Bridging the Terminology Gap using OWL:
A key problem in achieving interoperability is to be able to recognize that two pieces of data are talking about the same thing, even though different terminology is being used.
Example: Define the terms "Camera" and "SLR". State that SLRs are a type of Camera.

Here's how these two terms (classes) and their relationship is defined using the OWL vocabulary:

```
<owl:Class rdf:ID="Camera"/>
<owl:Class rdf:ID="SLR">
    <rdfs:subClassOf rdf:resource="#Camera"/>
</owl:Class>
```
**Scenario:**

- I am interested in purchasing a camera with a 75-300mm zoom lens size, that has an aperture of 4.5-5.6, and a shutter speed that ranges from 1/500 sec. to 1.0 sec.
- I launch my personal “web-bot” which crawls the web looking for web sites that can fulfill my request.
- Assume that there exists an OWL Camera Ontology, which the web-bot can “consult” upon its travels across the web.
The Web Bot finds this document at a website:

Is it relevant?
(Note: SLR = Single Lens Reflex)
I am interested in purchasing a camera with a 75-300mm zoom lens size, that has an aperture of 4.5-5.6, and a shutter speed that ranges from 1/500 sec. to 1.0 sec.

To determine if there is a match, these questions must be answered:
1. What's the relationship between "SLR" and "Camera"?
2. What's the relationship between "focal-length" and "size"?
3. What's the relationship between "f-stop" and "aperture"?
XML document is a match!

Web Bot (application)

hunts.xml

Semantic Definitions Separate from Application!

camera.owl

Semantic Definitions

"Relationship between Camera and SLR?"

"SLR is a type of Camera."

"Relationship between aperture and f-stop?"

"f-stop is synonymous with aperture."

"Relationship between size and focal-length?"

"focal-length is synonymous with size."

Institute for Software Technology and Interactive Systems
The example demonstrated how a web-bot was able to dynamically process an XML document from a Web site, despite the fact that the XML document used terminology different than was used to express the request. This interoperability was achieved by using the OWL Camera Ontology!

This example also demonstrated the *architectural design principle of cleanly separating the application code* (e.g., Web Bot) from the semantic definitions (e.g., Camera.owl).
An OWL Document is a Rich “Knowledge Bank” that your Applications can Tap Into

- Input Data
- Application
- OWL (knowledge bank)
- Output Data
Let's assume that the Camera application has access to an OWL Camera document:

**Terminology**

- The root element shows the Class of things being described, and the ID attribute identifies the instance of the Class.
- One or more Properties of the Class is provided.

This sample input data is following the RDF design pattern:
What Info in an OWL Document

- An OWL document contains this info:
  - **Class hierarchy info**: an OWL document defines class/subclass relationships.
  - **Synonym info**: an OWL document identifies equivalent classes and equivalent properties.
  - **Class association info**: an OWL document maps one or more classes to one or more classes, via a property (i.e., domain/range info).
  - **Property metadata info**: an OWL document contains a lot of metadata for properties.
  - **Class definition info**: an OWL document specifies the composition of classes.

- The following slides present examples to show how applications can use this info to enhance its robustness and extensibility.
Problem: design an application that filters inputs which do not contain Camera data.
Sample input:

```
<SLR rdf:ID="Olympus-OM-10">
  ...
</SLR>
```

This input contains Camera data and should not be filtered. If the application is designed to just look for `<Camera>` elements then it will incorrectly filter this input. However, by consulting the OWL Camera document (Camera.owl) it can dynamically discover that an SLR is a type of Camera:
Problem: design an application that filters inputs which do not contain Lens aperture data.
Sample input:

```
<Lens rdf:ID="Hasselblad_500V">
  <f-stop>…</f-stop>
</Lens>
```

This input contains Lens aperture data and should not be filtered. If the application is designed to just look for `<aperture>` elements then it will incorrectly filter this input. However, by consulting Camera.owl it can dynamically discover that f-stop is synonymous with aperture:

What do you know about f-stop?

f-stop and aperture are synonymous.
Problem: design an application that filters inputs which do not contain info about a Lens that is compatible with a Canon-EOS-1 body.

Sample input:

```
<Lens rdf:ID="Sigma_24mm_f2.8">
  ...
</Lens>
```

This input contains info about a Lens that is compatible with the Canon-EOS-1 body. If the application is designed to just look for Canon lenses then it will incorrectly filter this input. However, by consulting Camera.owl it can dynamically discover the compatible lenses: (assume a property, isCompatibleWith, which maps Canon-EOS-1 body to compatible lenses)
Using Property Metadata to Enhance your Applications

Problem: design an application that filters inputs which do not contain info about the Canon-Sure-Shot-Z155 (serialNum="ABCD").
Sample input:

```
<rdf:Description>
  <serialNum>ABCD</serialNum>
</rdf:Description>
```

This input contains Canon-Sure-Shot-Z155 data and should not be filtered. If the application is designed to just look for `<Camera rdf:ID="Canon-Sure-Shot-Z155">` then it will incorrectly filter this input. However, by consulting Camera.owl it can dynamically discover that the serialNum uniquely identifies Canon-Sure-Shot-Z155:

```
What do you know about serialNum?
serialNum is inverseFunctional.
```

Institute for Software Technology and Interactive Systems
OWL provides lots of metadata for properties:

- SymmetricProperty
- FunctionalProperty
- TransitiveProperty
- inverseOf
- InverseFunctionalProperty
- allValuesFrom
- someValuesFrom
- hasValue
- cardinality
- minCardinality
- maxCardinality

The ways that your application takes advantage of this metadata is limited only by your imagination!
Problem: design an application that filters inputs which do not contain info about MediumFormat cameras.

Sample input:

```xml
<rdf:Description rdf:ID="Hasselblad_500V">
  ...
</rdf:Description>
```

The MediumFormat class is defined in Camera.owl as an enumeration list. By consulting Camera.owl the application can retrieve the list, and thus dynamically determine that this input describes a MediumFormat camera:
Is it required that there be just one ontology for a domain?

Issue: How many versions of an ontology are allowed?  
Example: How many versions of a Camera ontology are allowed on the Web?

There are two cases:

1. Only a single Camera ontology is allowed - there is universal agreement on a Camera ontology.  
Comment: in a highly distributed, autonomous environment such as the Web, this is not realistic.

2. Every application has its own Camera ontology (ontology-per-application)  
Comment: the interoperability will broken down.

Here is the middle ground: The number of Camera ontologies is greater than one, but less than one-per-application.

The Evolution of a "Web of Camera Ontologies"

Consider this scenario: A community comes together to create a Camera ontology. Within that community there is perfect interoperability.

Independently, another community comes together to create another version of the Camera ontology. Within that community there is perfect interoperability, but across the two communities there is zero interoperability.
Is it required that there be just one ontology for a domain?

- Someone in the first community recognizes that a class in that community's Camera ontology is equivalent to a class in the other community's Camera ontology, so this equivalence is added to both ontologies (using `owl:equivalentClass`). Therefore, a bridge has been created between the two Camera ontologies, and now there can be interoperability between the two communities.

- Ultimately, all the different Camera ontologies get linked together, and there is global interoperability.

Now going the other direction (from macro to micro)...

- Within a community there may be subcommunities which refine the community's Camera ontology. Provided the subcommunity's ontology is linked to the community's ontology then there will be perfect interoperability.

- Therefore, there can be diversity in ontologies both at a micro level as well as at a macro level.
Upper-, Mid-level, Lower-Ontologies

- An upper-ontology defines very broad, universal Classes and properties
  - Example: Cyc Upper Ontology
    [http://www.opencyc.org](http://www.opencyc.org) (6,000 concepts)
- A mid-level ontology is an upper ontology for a specific domain
- A lower-ontology is an ontology for a specific domain, with specific Classes and properties.
- Example: Suppose that we create a mid-level ontology. We can merge into an upper-level ontology by defining our ontologies root class as a subClassOf a class in the upper-ontology.
Database = Ontology + Instances

<table>
<thead>
<tr>
<th>title</th>
<th>author</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>My Life and Times</td>
<td>Paul McCartney</td>
<td>June, 1998</td>
</tr>
<tr>
<td>Illusions</td>
<td>Richard Bach</td>
<td>1972</td>
</tr>
<tr>
<td>First and Last Freedom</td>
<td>J. Krishnamurti</td>
<td>1974</td>
</tr>
</tbody>
</table>

<?xml version="1.0"?>
<BookCatalogue>
  <title>My Life and Times</title>
  <author>Paul McCartney</author>
  <date>June, 1998</date>
</BookCatalogue>
Advantages of using OWL to define an Ontology:
- Extensible: much easier to add new properties. Contrast with a database - adding a new column may break a lot of applications (see example on next slide)
- Portable: much easier to move an OWL document than to move a database.

Advantages of using a Database to define an Ontology:
- Mature: the database technology has been around a long time and is very mature.
Protégé
Protégé Features

- Protégé is an extensible and customizable toolset for constructing ontologies and for developing applications that use these ontologies.

- Outstanding features:
  - Automatic generation of graphical-user interface, based on user-defined models, for acquiring domain instances.
  - Extensible knowledge model and architecture.
  - Possible embedding of standalone applications in Protégé knowledge engineering environment and vice versa.
  - Scalability to very large knowledge bases.
Storage Formats

- User edit and view ontologies in a manner that insulates them from the ultimate storage format.
- Ontologies may be read in from, written out to, and inter-converted between a large number of formats:
  - Relational database (ODBC)
  - UML
  - XML/XML Schema
  - RDF
  - Topic Maps
  - DAML+OIL
  - OWL
Role of Protégé in the Semantic Web

Semantic Web

Paradigms > Languages > Tools > Applications

- Paradigms:
  - Description Logics
  - Frame / OO

- Languages:
  - OWL
  - OKBC

- Tools:
  - ezOWL, etc.
  - Protégé
  - Protégé OWL Plugin

- Applications:
  - Decision Support
  - Classification
  - Reasoning
OWL Plugin

- Extension of Protégé to allow editing OWL ontologies
- Project started April 2003, based on ideas from previous projects (OilTab, RDF and DAML+OIL backends)
- Currently in: [http://protege.stanford.edu/plugins/owl](http://protege.stanford.edu/plugins/owl)
- Features:
  - Loading and saving OWL files
  - Graphical editors for class expressions
  - Access to description logics inference components such as classifiers (RACER)
  - Powerful platform for hooking in custom-tailored components.
Protégé/Properties (Slots)
Protégé/ezOWL Plugin

The image shows the user interface of Protégé/ezOWL Plugin, a tool for creating and managing ontologies. The interface is divided into several panels:

- ** asserted hierarchy panel**: Displays a tree structure of classes and properties.
- ** properties panel**: Contains a list of properties such as `Content`, `Day`, `DOW`, `EDate`, `ETime`, `Hour`, `Minute`, `Month`, `Year`.
- ** ezOWL panel**: Shows part of the ontology diagram with classes `Date`, `DaysOfWeek`, `Time`, `Events`, and properties `protege:valuesComputer`, `minCardinality`, `maxCardinality`, `Day`, `Month`, `Year`, `Minute`, `Hour`, `DOW`. The diagram illustrates the relationships and constraints between these entities.

The interface is designed to facilitate the creation and manipulation of ontologies, allowing users to define classes, properties, and relationships within an ontology.
Thank you very much