Importing Data into Protégé-OWL

Martin O’Connor
Stanford Center for Biomedical Informatics Research,
Stanford University
Importing Data into Protégé-OWL

- Most data are not stored in the form of ontology
- Goal: import and reason with data from external sources
- Several Protégé-OWL Plug-ins to support batch importation from relational database, spreadsheets, XML documents
- Ongoing research: dynamic importing
Importation Functionality

Databases
Spreadsheets
XML

OWL Ontology

Classes
Properties
Instances
Data Values
Protégé-OWL
Batch Importing Plug-ins

• DataMaster: relational importing

• SpreadSheetMaster: spreadsheet importing

• XMLMaster: XML importing
DataMaster Plug-in

• Imports relational schema and/or data

• Two import options:
  – Schema import only:
    • Schema as OWL classes
    • Schema as instances of Relational.OWL classes
  – Schema and content importation

• Supports any relational database with JDBC driver

• Java API for embedding in applications
Example Schema
Representing Schema as OWL Classes
Representing Schema in Relational.OWL Ontology
Representing Schema in Relational.OWL Ontology
Importing Schema and Content

- Imports relational schema and data
- Two schema representations:
  - Schema as OWL classes
  - Schema as instances of Relational.OWL classes
- Data stored as instances of either OWL classes or Relational.OWL instances
DataMaster Plug-in
SpreadSheetMaster Plug-in

- Protégé-OWL Plug-in for importing spreadsheet content into OWL ontologies
- Can import arbitrary Excel or CSV files
- Also, supports definition of mapping templates to import files of a particular structure
- Mappings saved as an OWL ontology
Example Source Excel Spreadsheet

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GOOG</td>
<td>34.55</td>
<td>23.22</td>
<td>23.44</td>
<td>43.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IBM</td>
<td>56.44</td>
<td>54.02</td>
<td>18.33</td>
<td>12.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Excel spreadsheet image with data shown)
Example Target Ontology
SpreadSheetMaster Screenshot

### Workbook

<table>
<thead>
<tr>
<th>Sheet1</th>
<th>Sheet2</th>
<th>Sheet3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Q1</td>
<td>Q2</td>
</tr>
<tr>
<td>2</td>
<td>GOOG</td>
<td>34.55</td>
</tr>
<tr>
<td>3</td>
<td>IBM</td>
<td>56.44</td>
</tr>
</tbody>
</table>

### Workbook File

C:\Development\Clados\xls\ShareDemo.xls

### Mappings Control

#### Trade

<table>
<thead>
<tr>
<th>Class name</th>
<th>Sheet name</th>
<th>Start column</th>
<th>Finish column</th>
<th>Start row</th>
<th>Finish row</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade</td>
<td>Sheet1</td>
<td>B</td>
<td>E</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Instance Property Mappings

<table>
<thead>
<tr>
<th>Property name</th>
<th>Property type</th>
<th>Property value</th>
<th>Value type</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasShare</td>
<td>Data</td>
<td>A$*</td>
<td>xsd:string</td>
</tr>
<tr>
<td>hasAmount</td>
<td>Data</td>
<td>$*$</td>
<td>xsd:float</td>
</tr>
<tr>
<td>hasQuarter</td>
<td>Data</td>
<td>$*$1</td>
<td>xsd:string</td>
</tr>
</tbody>
</table>
SpreadSheetMaster Features

- Can import content as classes, properties, instances, or data values
- Define complex mappings interactively with custom expression language
- Can save mappings and reuse them on documents with the same structure
- Java API for embedding in applications
- Available in a month
XMLMaster Plug-in

• Protégé-OWL Plug-in for importing XML documents into OWL ontologies
• Can import arbitrary XML documents
• Also supports definition of mappings templates to import XML documents of a particular type
• Mappings saved as an OWL ontology
• Java API for embedding in applications
• Available in a few months!
Dynamic Importing

• DataMaster, SpreadSheetMaster, and XMLMaster perform batch importation
• *Dynamically* importing content is desirable in some cases
• Entities are mapped to OWL axioms on demand
• Significant scalability advantages
Current Work:
Dynamic Importing Plugin

- Supports mappings to import relational, spreadsheet and XML data *on demand* when executing SWRL rules
- Mappings saved as an OWL ontology
- Available later this year!
Conclusion

• DataMaster: available Protégé 3.4
• SpreadSheetMaster: available Protégé 3.4 end of July
• XMLMaster: available Protégé 3.4 in a few months
• Dynamic importing tools: available Protégé 4 towards end of year
• Other tools will be ported to Protégé 4