Knowledge Engineering Tools and Semantic Translation

Andrzej Bialecki
WebGiro, Chief System Architect
<abial@webgiro.com>

ECIMF Project Group Meeting, Brussels, 16.10.2001
Knowledge Engineering Tools

- **Ontology Engineering (OE)**
  - Ontology creation, editing, comparing ...

- **Knowledge Acquisition (KA)**
  - From human experts
  - From machine-readable sources

- **Automated Reasoning (AR)**
  - Inference
  - Verification

- Often support multiple aspects
Semantic Translation and ontologies

**Ontologies**
- Shared conceptualization of a domain, “consensus view”
- Semantic content limited to the domain of interest

**Approximate re-classification**
- Semantic enrichment (needed for disambiguation)
- Upper-level ontologies (shared vocabularies)
Why KE Tools for ST?

- ST relies on the use of ontologies
  - Efficient ontology management needed
- Shared ontology design
  - Efficient ontology design support needed
- Validation and inference
  - Consistency checking and discovering non-obvious relations are crucial
- KE is a widely-researched topic
  - There are many supporting tools and models
  - BUT few of them support ontology mapping
Protégé 2000

- Features: OE, KA, AR
- Knowledge model: frame-based
  - Class, slot
  - Multiple inheritance
  - Template slots
- Modular and extensible
  - Plug-in system, with multitude of modules available (KA, inference, merging, WordNet...)
- Open Source (MPL)
Protégé Interface: Classes, slots

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Cardinality</th>
<th>Other Facets</th>
</tr>
</thead>
<tbody>
<tr>
<td>byname</td>
<td>String</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>current_job_title</td>
<td>String</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>date_hired</td>
<td>String</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>name</td>
<td>String</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>other_information</td>
<td>String</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>phone_number</td>
<td>String</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>responsible_for</td>
<td>Instance</td>
<td>multiple</td>
<td>classes=(Employee)</td>
</tr>
<tr>
<td>salary</td>
<td>Float</td>
<td>single</td>
<td></td>
</tr>
<tr>
<td>sections</td>
<td>Instance</td>
<td>multiple</td>
<td>classes=(Section)</td>
</tr>
</tbody>
</table>
Protégé Interface: OntoViz
Anchor-PROMPT: semantic matching

- Requires providing pairs of anchors
  - These can also be suggested based on heuristics (e.g. lexicographical similarity)
- Suggests matching terms by analyzing non-local context
  - Tests have shown ca. 60% correct guesses
Anchor-PROMPT: e-commerce scenario

- Two ontologies
  - O1: some well-known standard (e.g. ebXML)
  - O2: simple general ledger database tables
    - Tables -> classes
    - Fields -> slots

- User with limited ontology knowledge (typical for SME)
  - User can provide the most obvious matches
    - E.g. O1(Person) -> O2(Person)
  - Anchor-PROMPT suggests less obvious matches
    - E.g. O1(Address) -> O2(Location)

- Current implementation:
  - Only earlier version available (plain PROMPT)
  - Only ontology merging or sub-selecting, not mapping
Protégé Interface: PROMPT
Protégé Summary

- Very flexible and extensible
- Cross-platform, open source (MPL)
- Rich functionality
- Extensive API for model management
- Explicit support for semantic matching
- BUT: the knowledge model is difficult for non-experts
  - This could be hidden behind a special-purpose user interface module
Conzilla Concept Browser

- Features: OE, KA
- Cross platform, open source (GPL/MPL)
- Knowledge model: “neurons”
  - Associated with multilingual meta-data
  - Neuron types:
    - Class
    - Specialization (subclass-of [= isa in KE])
    - Association (related-to)
    - Aggregation (part-of)
    - Instantiation (is-a [≠ isa in KE!])
Conzilla Unified Language Modeling

- A UML profile, readable in natural language
- More precise and concise than traditional KE notation (isa, instanceOf)
Conzilla Interface

- Easily converted to/from natural language
- Easily accessible meta-data (multi-lingual)
- Easily accessible content
- Understandable for both business and technical people
Conzilla Summary

- **Unique context navigation**
  - Helps to keep track of the concept relationships

- **UML profile for diagrams**
  - UML is well-understood for a wider audience
  - ULM - diagrams easily converted to/from natural language

- **BUT:**
  - Currently uses its own XML format (RDF in works)
  - Difficult content editing
Conclusions: Protégé/Conzilla/PROMPT

Conzilla as a Protégé plug-in:
- Adds a non-expert interface to the KB
- Easier to understand graph. visualization of the ontology
- Ontology can be exported to many formats
- All other Protégé plug-ins become available for additional processing
- Advanced Protégé interface available for experts

Anchor-PROMPT can help less experienced users

All code is under open source license (MPL)
Further information

❖ Protégé website
   ■ http://protege.stanford.edu

❖ Anchor-PROMPT

❖ Conzilla
   ■ http://www.conzilla.org

❖ ECIMF Project Information Center
   ■ http://www.ecimf.org