Two Protégé plug-ins for supporting document-based ontology engineering and ontological annotation at document-level - aka Domain Modeling Tool (DMT)

Viktoria Pammer, Peter Scheir, Stefanie Lindstaedt

aposlle – New ways ...

... to work, learn and collaborate!
Outline

- Motivation
- Document-based ontology-engineering
- Ontology-based document annotation
- Intermediate résumé
- Conclusion
Introduction

**APOSdle**

- System supporting work-integrated learning
- "**Knows**" about users, tasks, task and application domain knowledge, competencies, resources
- "**Understands**" resources
- [http://www.aposdle.org](http://www.aposdle.org)
Introduction cont’d

- **Where do we have knowledge?**
  - Heads of experts
  - Documents (FAQ, guidelines, project reports, MATLAB simulations etc.)

- **In what form do we need it?**
  - Formal language (OWL)

- **How do we make use of formal knowledge?**
  - Annotate resources and experts
  - Some other things…
Step 0
Scope & Boundaries and Resources Collection

Step 1
Knowledge elicitation from digital resources

Knowledge elicitation from Domain Experts

Step 1a
Step 1b

Step 2
Informal Modelling

Step 3
Informal Models Integration & Revision

Step 4
From Informal to Formal

Step 5
Formal Models Integration

Step 6
Formal Models Validation

Credits go to Marco Rospocher (FBK) for this picture
Document-based ontology engineering

- Given (textual) documents about some domain, build an ontology
  - The principle is: automatic extraction of a formal knowledge model from non-formal information.
  - Is full automation possible?

- Question of technology: extract the same from multimedia data, from databases etc.
DMT: Discovery Tab

- Cluster documents
- Extract terms (and rank them)
- Group terms according to synonymity (WordNet)
- Export extracted terms to textfile
- Create class and property from term
Ontology-based document annotation

- **Annotate resources**

- **Use elements of a (formal) domain model for annotation**
  - Advantages over “tagging”: controlled vocabulary (facilitating automatic classification), navigation along ontology possible,
  - Disadvantage: overhead for modeling
DMT: Annotation Tab

- Manually annotate documents with classes and instances
  - Store annotations
  - Annotations are in OWL Full

- Classify new documents
  - Annotations are suggested with a weight, user needs to accept before these annotations are stored
Intermediate résumé + conclusion

- First-cut evaluation:
  - Users rate convenience-functionality (sorting lists, searching in terms etc.) very high
  - Quality of automation (term extraction, clustering, classification) is crucial for non-expert users
  - Informal modeling needs to be emphasized

- A number of standard text-processing methodologies grouped together within one tool
- Embedded in a standard ontology development tool
Thank you!

- [http://www.know-center.at/dmt](http://www.know-center.at/dmt)
- We welcome feedback!