Enterprise Vocabulary Development in Protege/OWL: Workflow and Concept History Requirements

Sherri de Coronado
Gilberto Fragoso

Protégé Workshop – Jul 8, 2004
Topics

• Background
• NCI Thesaurus conversion to OWL
• Requirements for Using Protégé-OWL for NCI Thesaurus
• Progress / Pilot Testing
NCI EVS

• Services and resources addressing NCI needs for controlled vocabulary
  http://ncicb.nci.nih.gov/core/EVS

• Goal: Integration by Meaning

• Collaboration between NCI OC and NCICB
  – Cancer Information Products and Systems (PDQ and Cancer.gov)
  – caCORE and Community portals
NCICB builds on EVS via caCORE Infrastructure

https://ncicb.nci.nih.gov/core
NCI Thesaurus

- Public domain, open content license
- Broad coverage of cancer domain
  - Neoplastic disease, Findings and Abnormalities, Anatomic Structures, Agents, Cancer-related genes, Gene products, etc.
- DL based using Apelon’s Ontylog
- 34,000+ “Concepts”
  - 20 hierarchies, 19 kinds
  - “Roles” establish semantic relationships between Concepts
  - “Properties” state facts about Concept
- Concept history
Ontylog to OWL Conversion

• Why OWL Lite for the conversion?
  – *To make it available in a non-proprietary form*
  – *To enable a wider audience to use it.*
  – *Current Thesaurus has fairly simple semantic constructs*
Mapping the Semantics

- Kinds and Concepts modeled as Classes
- Ontylog Role becomes ObjectProperty with Domain and Range (restrictions)
- Ontylog Property becomes AnnotationProperty
- Some and All translated as SomeValuesFrom and AllValuesFrom
Requirements for Using Protégé-OWL

• Concept History
• Search Capabilities
• Various Edit Actions / User Interface
• Workflow Management Functions
• Vocabulary Server (DTS or something new?)
Concept History Issues

• Certain editing actions result in retirement of Thesaurus codes
  – Merge, Split, Retirement

• Dependent applications/users require a mechanism to retrieve data coded with Thesaurus codes that have been retired

• Tracking complex edit actions in History allows dependent apps/users to query for replacement codes
Search Capabilities

• Must operate on various term-containing properties, not just on class names
  – *Good search capability critical for users and editors*
  – *Search on terms in annotation properties*

• Configurable, e.g. for default settings
Edit Actions / User Interface

• Support various editing actions
  – Merge
  – Split
  – Pre-retirements (by editor)
  – Retirement (by manager)
Split Edit Action

• Generates a new class
  – *History must record an association between the split and the new class*

• Properties and subclasses must be reviewed and resolved between the new and existing classes

• References to existing class must be reviewed and edited if necessary

• Must have GUI support
"Split" GUI Panel
New Class in Tree
Merge Edit Action

• Existing class is merged into another and retired
  – *History must record a retirement action, and an association between the surviving and the retired class*

• Properties must be copied, properties of retired class must be recorded (AnnotationProperty), subclasses must be moved to surviving class, retired class must be re-treed

• References to retired class must be reviewed and edited if necessary

• Must have GUI support
Merge Window
Select Surviving Class, Drop into Rightmost Pane
Retirement Actions

• Editors flag class for pre-retirement
  – Review and remove/modify restrictions and subclasses
  – State is annotated: super and subclasses, restrictions, references
  – References to class eliminated
  – Class is re-treed to holding bin, remaining subclasses re-treed under class' parent

• Manager confirms retirement
  – Class is re-treed to retirement bin
  – No programmatic Undo support
  – History records the retirement action, and associations to the class' parent classes

• GUI support for pre- and retirement
Pre-Retirement GUI

Subclasses

Restrictions
Workflow Management Needs

- Worklist assignments by manager and tracking of worklist items by editors
- Assignment of editing/review privileges
- Locking and unlocking of database (or server) for editing
- Review and consolidation of editing changes by manager
- Generation of reports by manager or editors
Other Workflow Needs

• Import Changesets by Manager and export Changesets by Editor (maybe)
• Export of database “Baseline” by manager
  – Development or Release baselines
  – Release export results in auto history export
• Configuration/constraints of environment
• Backup and Restore of database to archive by manager
Data Handling Issues

• Changed items should be flagged for review
• Consolidation/conflict resolution step involves accepting or rejecting changes to concepts/classes made by editors
• Class-instance deletion is restricted
• All edit actions processed in parallel for history
Progress/ Pilot Testing

• NCI Protégé/OWL extension in progress
  – NCIOWLClsesTab to support workflow/history as shown

• Pilot to Evaluate Protégé-OWL for editing and semantic capabilities
  – 2-3 months: Kevric, NCI, Stanford, Uvic
EVS Team

EVS

NCI OC – oncology, pathology, pharmacy
  Margaret Haber
  Larry Wright
NCI CB – biology, operations
  Sherri de Coronado
  Gilberto Fragoso
  Frank Hartel

Apelon, Inc.  Northrop Grumman, Inc.
Aspen, Inc.    Kevric Corporation
Jim Oberthaler  SAIC
Stanford Medical Informatics