Community-based ontology development, alignment, and evaluation

Natasha Noy
Stanford Center for Biomedical Informatics Research
Stanford University
Community-based Ontology ... Everything

Development and Evolution

Evaluation and Reuse

Mapping and Alignment
How Large Can the “Community” be?

Can be just a few distributed users

Or anyone in the world
Collaborative Ontology Development

- **Collaborative**
  - Several users contribute to a single developing ontology
  - There are mechanisms to carry out discussions and to reach consensus

- **Ontologies**
  - From simple taxonomies
  - To expressive OWL ontologies
Use Cases in Biomedical Domain

- Gene Ontology (GO)
- NCI Thesaurus
- BiomedGT
- OBI, BIRNLex
The Gene Ontology

Terminology for consistent description of gene products

- 3 full-time curators have access to edit GO
- Anyone in the community can submit an issue or request
The NCI Thesaurus

Reference ontology for cancer biology, translational science, and clinical oncology

- ~20 full-time editors making changes
- Changes are not immediately visible
- 1 “lead editor” who approves the changes, and assigns new tasks
### BiomedGT

An “open” version of the NCI Thesaurus

<table>
<thead>
<tr>
<th>Assign subset of ontology</th>
<th>Create proposal</th>
<th>Assign recommendations</th>
<th>Implement changes</th>
<th>Notify of new changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role: Alpha SME</td>
<td>Role: SME</td>
<td>Role: Alpha curator</td>
<td>Role: Curator</td>
<td>Role: System</td>
</tr>
<tr>
<td>Alpha SME assigns subset of ontology to SME</td>
<td>SME creates a new proposal for a change of content</td>
<td>Alpha curator will assign recommendations to curators</td>
<td>Curator will implement the changes in the terminology</td>
<td>The system will send notifications to interested parties that the change has been integrated</td>
</tr>
</tbody>
</table>

- Several different roles
- Capabilities depend on the user’s role
- Anyone in the community can contribute suggestions; small number of curators
Other Biomedical Projects

- OBI, BIRNLex, RadLex
  - tightly-knit community of developers (20-30)
  - most actively participate in discussions
  - 1 or 2 editors make changes to the ontology
Dimensions of Collaborative Workflows

- **Ontology size**
  - from 100s to 10,000s of concepts

- **Size of the community**
  - Contributors (in some form): from 2-3 to dozens
  - Editors: from 1-2 to 20

- **Control mechanisms**
  - Variety of roles
  - Gatekeepers, etc.
  - Client-server editing

- **Discussion tools**
  - mailing lists, message boards
  - face-to-face meetings, telecons

- **Synchronization and editing mechanisms**
  - CVS, SVN
Tool Requirements

• Tools for **discussion and reaching consensus**
  • annotate components and, maybe, changes
  • have as an integral part of the development process

• **Context** for discussions on modeling decisions

• **Record of changes** and associated discussions and controversies

• **Provenance and trust**
  • support concept histories
  • have ways establish trust and credibility

• **Personalized views** of an ontology
  • based on user’s role and tasks
  • based on user’s level of expertise
  • based on user’s trust network
Tool Requirements (cont’d)

• Support for **personal and shared** spaces
• **Access control**
  • fine-grained control for editing and viewing rights
• **User roles**
• **Flexible workflow support**
  • configurable workflows
  • workflow-execution coupled with ontology development
Bringing Collaboration to Protégé

• Protégé
  • is an open-source ontology editor
  • developed at Stanford, in collaboration with University of Manchester
  • with more than 100,000 registered users
• has dozens of plugins for
  • visualization
  • inference
  • import and export
  • natural-language processing
• ... (things we don’t know exist)
Collaborative Protégé

- An extension of the Protégé to support collaborative development
Features of Collaborative Protégé

• Support for:
  • annotating ontology components and changes in the ontology
  • discussion threads
  • proposals and voting
  • searching and filtering
  • defining users, groups, policies

• Distributed with Protégé installation

http://protege.stanford.edu/doc/collab-protege/
Annotations

- Annotations are linked to a specific ontology component:
  - class
  - property
  - individual
  - ontology as a whole
  - ontology change
  - ... others

- There are different types of annotations:
  - question
  - comment
  - proposal

- Users may filter and search annotations based on different criteria
Collaborative Protégé Interface

Collaborative Panel

Collaborative Tabs

Annotations

has annotations

Annotation details
Chat

Users currently on-line

Hyperlinks to the entities in the ontology
Architecture

Domain Ontology

Protégé Client

Instances of the Change and Annotation Ontology

annotate
record changes
create annotations

make changes

instances of

annotate
Changes and Annotations

Change
applyTo
author
created

Class_Change

Restriction_Added
Superclass_Added

Property_Change

Individual_Change

AnnotatableThing

Annotation
title
author
created
modified

annotates
assoc_annotations

Proposal
Explanation
Comment
Example
Advice

Recommendations:
- Proposal
- Explanation
- Comment
- Example
- Advice
But wait, there is more...

Discussions and proposals is about Changes

Ontology components/axioms is about

is about applies to approve has privilege

create proposal for Workflows

Roles
Support For Flexible Workflows

Goal:

- automatically generate tools that support custom-tailored workflows
- cover a wide range of existing workflows
Custom-Tailored Workflows

Workflow instantiation

Workflow Engine

Execution Workflow

State manager

Logging

Action adapter

Action adapter

Action adapter

Plug-ins
Workflow Ontology

- AnnotatableThing
  - OntologyComponent
  - Annotation
  - Proposal
  - Task
  - SequenceActivity
  - ListenActivity
  - DelayActivity
  - OnTaskCompletedActivity
- Workflow
  - AssociatedWorkflows
  - AppliesTo
  - InitiationForm
  - RootActivity
- Activity
  - TransitionActivity
  - State
    - Branch
  - ControlState
  - ChoiceState
  - ForState
  - EmailUserActivity
  - CreateTaskActivity
  - EventDrivenActivity
  - CompositeActivity
  - AtomicActivity
Community-based Ontology ... Everything

- Development and Evolution
- Evaluation and Reuse
- Mapping and Alignment
NCBO BioPortal

• The National Center for Biomedical Ontology (http://bioontology.org) is developing BioPortal, an open-source repository of ontologies, terminologies, and thesauri of importance in biomedicine.

• An early version of BioPortal is accessible at http://bioportal.bioontology.org. An alpha version of the next release is at http://alpha.bioontology.org/

• Users can access the BioPortal content interactively via Web browsers or programmatically via Web services.
The BioPortal Ontology Repository

- Open repository of ontologies in biomedicine
- Each ontology is described by a set of metadata
- Ontologies in different formats
  - OWL, RDF(S)
  - OBO
  - Protégé frames
- BioPortal technology
  - open-source
  - domain-independent
Ontology Navigation in BioPortal
Major Function: Ontology Assessment

Which ontology is appropriate for my task?

Sources of information for the answer:

- Ontology metadata
  - *usually provided by authors*
- Computable metrics
  - *can be provided by the tools*
- Community-based evaluation
  - *provided by other users of the ontologies*
Ontology Metadata

- Provides answers to pertinent questions about the content and provenance of the ontology in the library:
  - What is the domain covered by an ontology?
  - What are the key classes and concepts?
  - Who developed the ontology?
  - What is the policy for maintenance and distribution?
  - What is the format of the ontology (syntax, language, tools used to build it, etc.)
Computable ontology metrics

• Logical consistency checking
  • e.g., use a reasoner to determine if an ontology is consistent

• Structural consistency based on meta-properties
  • e.g., OntoClean

• Rules of thumbs and heuristics

• Statistical information
  • number of classes and properties
  • connectedness, fan-out, etc.
  • cycles
Not all useful metrics are computable

- Many aspects of ontology quality are subjective
- The most useful information for the user selecting an ontology:
  - *who used an ontology for a similar task and how well did it work?*
Some Ontology Metrics are Subjective

What is a “good” feature in some setting, can be a “bad” feature in another setting

- High level of axiomatization:
  - good if you want to perform reasoning
  - can be bad because of the high computational and cognitive cost if you don’t need the axioms
- Organizing anatomy concepts based primarily on their structure rather than function
  - can be good if you need to understand which organs a wound goes through
  - not appropriate if you need to understand spread of disease
Community-Based Evaluation

Which ontology from the library is appropriate for my task?

- The only people who know the answer to these questions are:
  - (maybe) ontology authors
  - other users of the ontology
- Allow users to provide ratings for ontologies
Open-Rating Systems

• Open publishing system
  • anyone can publish content
    • (Semantic) Web is inherently an open publishing system
    • many ontology libraries allow contributions by anyone

• Closed Rating system
  • only a group of “editors” can provide ratings
    • Open directory
    • Yahoo! directory

• Open Rating system
  • anyone can publish reviews and ratings
    • Amazon reviews
Could have been great, June 6, 2008
By C. Schumacher - See all my reviews

There is a great need for a book like this, and the authors are well-credentialed, but unfortunately, this book is more like an introduction than a craftsman's resource. It reads like a first draft with typos, belabored repetitive text, and some odd examples. There is almost no discussion of internationalization, security, performance, or tricky basic types such as dates, times, or currency. More discussion of SPARQL would have been useful, as well as discussion of what should go in the model vs. what should be queried out. The chapter "Good and Bad Modeling Practices" was a particular let-down. A sharp, thoughtful, deeper book on this topic would really help the field.

Was this review helpful to you? Yes No

Must read for all who want to get into the semantic web, June 5, 2008
By Henry Story "bblfish" (MPK20) - See all my reviews

Finally a book that explains the key aspects of the Semantic Web in easy to understand language, with well thought out examples and taking a step by step approach to this eminently stackable technology. No RDF/XML in the whole book! This will save all who get started on the semantic web years of work hunting down specs, and will give them the right initial intuitions.

A God send! Buy one for yourself and one for your friends too :-)

Was this review helpful to you? Yes No
Applying Open Ratings to Ontologies

Project Name: RadLex
Description: RadLex is a controlled terminology for radiology.

Institution: RSNA
People:
Home Page: http://radlex.org

Ontologies Used

RadLex

Degree Of Formality
Documentation And Support
Usability
Domain Coverage
Correctness
Quality Of Content

View Ontology Metadata

Anatomical entity
Non-physical anatomical entity
Physical anatomical entity
Immaterial physical anatomical entity
Material anatomical entity
Attribute entity
Dimensional entity

Physical anatomical entity

Visualization Details Marginal Notes Mappings Resources Annotations

Comment: Wrong superclass in Class/Details
ngriff at 01/31/08 17:09
The superclass for this class in the Class detail is wrong.

Comment:
ngriff at 01/31/08 17:10

New Thread
Reviewers Provide

- General review and rating
- **Usage** information
  - Which applications have successfully used the ontology?
  - What problems were encountered?
- **Coverage**
  - Does it cover the domain properly?
  - Are there major gaps?
  - Are some parts developed better than others?
- **Concept-specific comments**
  - Are there problems with specific concepts?
  - What alternative definitions should be used?
Conflicting Sources of Metadata

- Authors and users can contradict one another
  - Quality of documentation?
  - References (e.g., positive and negative analyses of the ontology)
  - ...
- Metadata schema must enable diversity of views on some metadata values
Representing Reviews and Ratings

Looks familiar?

- Discussions and proposals
  - is about
  - applies to
  - create proposal for

- Changes
  - is about
  - approve
  - has privilege

- Ontology components/axioms

- Workflows

- Roles
Community-based Ontology ... Everything

Development and Evolution

Mapping and Alignment

Evaluation and Reuse
Multiple and Overlapping Ontologies

- Ontologies, vocabularies, and terminologies will inevitably overlap in coverage
- An ontology library can provide declarative mappings
  - found by the tools (efficient, but far from perfect)
  - specified by users (low throughput, but better quality)

“Basically, we’re all trying to say the same thing.”
Mappings in BioPortal

Sural Artery

Mapping To | Source | Mapped By | Mapped On
--- | --- | --- | ---
external sural artery (Mouse adult gross anatomy) | NCICB | TerryHayamizu | 04/23/08
sural artery (Mouse adult gross anatomy) | NLM | SongmacZhang | 04/23/08
external sural artery (Mouse adult gross anatomy) | NLM | SongmacZhang | 04/23/08
superficial sural artery (Mouse adult gross anatomy) | NLM | SongmacZhang | 04/23/08
Creating Point-to-Point Mappings

Point-to-Point Mapping from Sural Artery to:

Ontology: Galen
Search: artery

Preview of SupraduodenalArtery.
### Analyzing Mappings

<table>
<thead>
<tr>
<th>Concept</th>
<th>Maps To</th>
</tr>
</thead>
</table>
| NCI Thesaurus : Pelvic Bone | Mouse adult gross anatomy : pelvic bone (1)  
Mapped By TerryHayamizu |
| NCI Thesaurus : Sural Artery | Mouse adult gross anatomy : external sural artery (2)  
Mapped By TerryHayamizu, SongmaoZhang |
| NCI Thesaurus : Spermatic Vein | Mouse adult gross anatomy : spermatic vein (2)  
Mapped By TerryHayamizu, SongmaoZhang |
Functionality for Mapping Support

- Enable users to
  - **upload** bulk mappings in a specified format
  - **download** mappings based on a selection criteria
  - **define** point-to-point mappings interactively
  - **comment** on mappings created by others
  - **refine** and **discuss** existing mappings
Mapping Metadata

• yes, metadata again...
• Users require a comprehensive set of metadata to augment mappings
  • mapping relationship
  • provenance (who created the mapping and when)
  • discussion and comments
  • application context
  • mapping dependency
  • algorithm used to create the mapping (configuration, parameters, etc.)
  • external references
Representing mappings

Mapping Ontology

Class:
One_To_One_Mapping
  source: URI
  target: URI
  relationship: URI
  metadata: Mapping_Metadata

Class:
Mapping_Metadata
  author: User
  created: Date
  dependency: One_to_one_mapping
  evolutionary_evidence: String

Instance: MM_456543
  author: natasha
  created: 3/24/08
  metadata: MM_456543

Mapping Instances

Instance: MP_01234
  source: http://ontology1.org/v1#Heart
  target: http://ontology1.org/v2.1#Heart
  relationship: http://mappingRelationships.org/v1.0#similarTo
  metadata: MM_456543
Current Mappings in BP

- More than 30,000 mappings
  - created manually
    - as part of concept definition (OBO xref, UMLS)
    - as a mapping (NCI Mouse-Human anatomy)
  - created automatically
    - Using algorithms such as Prompt
Mappings as the Product of Community Contribution

- Mappings can contradict each other
  - application context may be different
  - trust is the key again
- Users can use the BioPortal framework to reach consensus on the mappings
Use of Mapping Repository

- Source of data for automatic algorithm
  - machine learning
  - algorithms that need a priori alignment
- Accessible through web services
  - can be used in other applications
- Used for annotating and browsing resources through ontology elements
- Use for finding “important” ontologies:
  - If everyone maps to NCI Thesaurus, it must be important
Community-based Ontology ... Everything

Development and Evolution

Evaluation and Reuse

Mapping and Alignment

Are we there yet?
Research Challenges

• If we build it, will they come?
• How do we encourage users?
  • it is a paradigm shift, in some sense
  • so far, our users are asking for these features
    • collaborative ontology development is hot!
    • community-based mappings and evaluation is still new
One of the key issues: Trust

- Aggregation
- Meta-rankings: rating the raters
- Personalized and filtered views of the system: Web of Trust
- Online version of “word of mouth”
- Topic-specific trust

"On the Internet, nobody knows you're a dog."
Collaborative Workflows

• How do we develop a representation that is comprehensive enough to capture a wide variety of collaborative workflows?
• How do we make it simple and usable?
• What are the reusable workflow modules? Can we have a simple wizard that instantiates a workflow description?
Dynamics of Collaborative Ontology Development

- Properties and dynamics of the social networks that form as the result of collaborative ontology editing?
- What are the different types of users based on their activity in the editing process (e.g., those who lead the discussions vs those who do most of the editing)?
- What are the characteristics of the concepts that are discussed most actively?
- Do tools such as Collaborative Protégé make collaboration more efficient or hinder it?
Personalized Views of Ontologies

- How do we represent and compute personalized views based on the areas of interest?
- How do you enable local editing of ontologies with subsequent integration in the ontology being developed by others?
Maintaining Metadata through Ontology Versioning

Potentially, any part of the *description* can change: author, language, domain, ...

Mappings can become invalid, too...

Reviews may no longer apply...
Acknowledgements

• Mark Musen
• Tania Tudorache
• Abraham Sebastian
• Nick Griffith
• Sherri de Coronado
• Tim Redmond

• Jennifer Vendetti
• Nigam Shah
• Benjamin Dai
• Michael Dorf
• Harith Alani
• R.V. Guha
• Holger Lewen