Interactive Views for Navigating Ontologies and Data at the National Center for Biomedical Ontology

Margaret-Anne Storey
University of Victoria

Overview
Part 1:
- National Center for Biomedical Ontology
- What are ontologies and why use them?
  - Key challenges

Part 2:
- Our projects
  - Degree of interest visualizations with Diamond
  - Flexible and generic visualizations for Jambalaya and BioPortal
  - Visualizing ontology alignments: CogZ
  - Exploring clinical trials
- Future research directions

Participants in NCBO
- Stanford: Tools for ontology alignment, indexing, and management
- Lawrence–Berkeley Labs: Tools to use ontologies for data annotation
- Mayo Clinic: Tools for accessing large controlled terminologies
- Victoria: Tools for ontology visualization
- University at Buffalo: Dissemination of best practices for ontology engineering

Acknowledgements:
Chris Callendar, Tricia d’Entremont, Sean Falconer and Maleh Hernandez – The Chisel Group, UVic
Mark Musen, Principle Investigator NCBO – Stanford University
Ida Sim, Driving Biological Project Lead, San Francisco University

National Center for Biomedical Ontology

Goal: develop innovative technology and methods that allow scientists to record, manage, and disseminate biomedical information and knowledge in both human readable and machine-processable form

NCBO Driving Biological Projects
- Flybase (Cambridge UK)
- ZebraFish (ZFin project, Oregon)
- HIV/AIDS Clinical Trial Bank
  (University of California at San Francisco)
What is an Ontology?

- The study of being
- A discipline co-opted by computer science to enable the explicit specification of
  - Entities
  - Properties and attributes of entities
  - Relations between entities
- A theory that provides a common vocabulary for an application domain

Why develop an ontology?

- To share a common understanding of the entities in a given domain
  - among people
  - among software agents
  - between people and software
- To enable reuse of data and information
  - avoid re-invention of the wheel
  - introduce standards to allow interoperability and automatic reasoning
- To create communities of researchers

Ontologies are pervasive...

A successful example: Foundational Model of Anatomy

- Long-term project at University of Washington to create a comprehensive ontology of human anatomy
  - 72K concepts, 1.9M relationships
  - One of the largest and best developed ontologies in biomedicine

Top level of the foundational model of anatomy
Ontologies are needed in electronic form

- Ontology contents can be processed and interpreted by computers
- Interactive tools can assist developers in ontology authoring

Protégé Demo

Ontologies enable large-scale science

Ontologies are currently at the center of two major activities in biomedical research

Structured representation of biomedicine:
- Communities of researchers are creating and maintaining biomedical ontologies to represent different types of entities and relations to describe biomedicine (ontology content curation)
- Annotation: biomedical experimentalists use ontologies to summarize and describe their results in a structured manner, enabling:
  - Integration of their data with other researchers’ results
  - Cross-species analyses, trial meta-analyses

Challenges

- But many ontology builders are not very good philosophers!
  - Nearly always, ontologies are created to address pressing practical needs
  - Most ontologies are relatively small, built and refined by small groups
  - Success rests on individual artisans, rather than on standard operating procedures
  - Developers often lack training in formal modeling and languages used may be limited

For example…

The International Classification of Diseases

- An enumeration of diseases that forms the basis for all medical claims and reimbursements in the United States
- A "legacy" terminology that has its roots in 19th century epidemiology
- Created initially by biostatisticians with a pressing need to compare death statistics in different European countries

The International Classification of Diseases: An excerpt…

- 724 Unspecified disorders of the back
  - 724.0 Spinal stenosis, other than cervical
  - 724.00 Spinal stenosis, unspecified region
  - 724.01 Spinal stenosis, thoracic region
  - 724.02 Spinal stenosis, lumbar region
  - 724.09 Spinal stenosis, other
  - 724.1 Pain in thoracic spine
  - 724.2 Lumbago
  - 724.3 Sciatica
  - 724.4 Thoracic or lumbosacral neuritis
  - 724.5 Backache, unspecified
  - 724.6 Disorders of sacrum
  - 724.7 Disorders of coccyx
  - 724.70 Unspecified disorder of coccyx
  - 724.71 Hypermobility of coccyx
  - 724.71 Coccygodynia
  - 724.8 Other symptoms referable to back
  - 724.9 Other unspecified back disorders

ICD9 (1977): A Handful of Codes for Traffic Accidents

- 820-829 Supplementary classification of external causes of injury
  - 820 Motor vehicle traffic accidents
    - 820.1 Motor vehicle traffic accident involving other off-road vehicle
    - 820.2 Other motor vehicle traffic accident involving other off-road vehicle
    - 820.3 Other motor vehicle traffic accident involving pedestrian
    - 820.4 Other motor vehicle traffic accident involving bicycle
    - 820.5 Other motor vehicle traffic accident involving motorcycle
    - 820.6 Other motor vehicle traffic accident involving aircraft
    - 820.7 Other motor vehicle traffic accident involving aircraft
  - 821-829 Other traffic accidents
    - 821-829.1-829.99 Other traffic accidents

ICD10 (1999): 587 codes for such accidents

V31.22 Occupant of three-wheeled motor vehicle injured in collision with pedal cycle, person on outside of vehicle, nontraffic accident, while working for income
W65.40 Drowning and submersion while in bath-tub, street and highway, while engaged in sports activity
X35.44 Victim of volcanic eruption, street and highway, while resting, sleeping, eating or engaging in other vital activities

ICD is used for lots of (too many?) things!
- ICD is used to code all patient encounters with the health-care system for:
  - Billing and reimbursement
  - Institutional planning
  - Disease surveillance and public health
  - Quality assurance
  - Economic modeling
- ICD was never intended to make the distinctions relevant to all these tasks!

If real ontologists could build the ICD from scratch …
- Diseases would be organized with well-defined relationships
- Diseases would be associated with computer understandable definitions
- There would be well-defined rules for ensuring that descriptions are sensible
- There would be well-defined mechanisms for creating user-specific views of the ICD
- There would be integration with bioinformatics resources that describe the molecular underpinnings of disease

Summary of challenges
- Ontologies are used in many areas of science:
  - e.g. Gene Ontology -- new genes discovered daily
- Lack of agreement on meaning of familiar terms
- Ontologies need to be peer reviewed and updated through consensus
- Data annotations also have to be updated as ontologies change!
- Reality is that many ontologies will have to coexist -- need mappings between the ontologies
- BUT! the ontologies, mappings and data annotations will also have to evolve...
- Issues of collaboration and trust...

Goals of the Center Revisited

Unify the divergent and isolated efforts in ontology development by promoting high quality open-source, standards-based tools to create, manage, and use ontologies

Create new software tools so that scientists can use ontologies to annotate and analyze biomedical data

Disseminate the tools and resources of the Center and to identify, evaluate, and communicate best practices of ontology development to the biomedical community

NCBO plans to offer technologies
- To help build and extend ontologies
- To locate ontologies and to relate them to one another
- To visualize relationships and to aid understanding
- To facilitate evaluation and annotation of ontologies
Our goal: Provide cognitive support for ontology developers and users through visual and intelligent user interfaces

Subprojects:
- DIaMOND—Degree of Interest Modeling for Ontology Navigation and Development
  - http://www.thechiselgroup.org/diamond
- CogZ: Cognitive support for ontology alignment
- Jambalaya: visualization in Protégé

Motivation
- Navigating ontologies can be tedious...
  - Long scrolling lists, expanding/collapsing nodes
  - Large number of irrelevant elements occlude relevant information
- Users often don’t know where to start when navigating an unfamiliar ontology
  - Might appreciate “worn paths”
DiMOND (plug-in)

- Three levels of interest
  - **Interesting**
    - Has been interacted with such that the DOI value exceeds a threshold value
  - **Uninteresting**
    - DOI value falls below the threshold value
  - **Landmark**: Hub concept
    - Manually specified by user
    - DOI value exceeds a threshold value
  - Lightweight, easily reversible focus techniques
  - **Consistent** with existing, familiar Protégé views

Diamond Demo

Diamond's Future Work

- Currently being evaluated in **user studies**
- **Sharing** DOI among users (many requests)
- **Role and Task-based** DOI calculations
- Use of instance data and annotated data to supplement DOI calculations

Integrate and evaluate Diamond approach in **BioPortal**

Jambalaya

Visualization of ontologies to support navigation

Jambalaya Demo

Visualization and Jambalaya

- Why are visualizations important?
  - Ontologies can be extremely large and complex, difficult to navigate using just trees and lists
  - Visualizations help users understand ontologies by showing overviews and hub concepts
- Other visualization approaches:
  - OntoViz, OWLViz, TGViz, ezOWL
  - IsaViz: visual RDF editor
  - Many graph-like visualization tools!
- What is Jambalaya?
  - Protégé Plug-in built on top of SHRiMP

Main Features in Jambalaya

- Classes and instances/individuals are represented by nodes in the graph
- Slots or properties are represented by arcs between nodes
• High level edges (lifted)

• Inconsistencies "pop-out" in some views
Different Views
- Nested View (Default)
- Flat Views
  - Class Tree
  - Class & Instance Tree
  - Domain & Range (OWL)
- TreeMap View
- Query View
- Filmstrip
- Customized views

FlexViz
- Currently implemented as a Java Applet
- Integration of key ideas from Jambalaya with BioPortal
  - In progress... determining required features

CogZ
- Cognitive Support and Visualization for Human-Guided Mapping Systems

Ontology alignment
- Mapping terms from one ontology to another, preserving structural relationships
- Research emphasis mostly on the algorithm

Ontology alignment continued
- Cognitively challenging
- The user must:
  - Make decisions about mapping candidates
  - Supply custom mappings that were missed
  - Understand the domains and structure of both ontologies
  - Remember decisions that were made
- Can be made easier through cognitive support

Visualization plug-in for Prompt
- **Prompt:** framework to support ontology alignment
- Sean Falconer’s research:
  - Added a **plug-in mechanism** and **perspectives** to Prompt to support other algorithms and visualizations
  - CogZ: Integrating various **visualizations** and **filters** to support the user's decision making process

**CogZ Demo**
Future work in visualizing alignments

- Carry out studies to investigate the cognitive processes involved during ontology mapping
- Develop a model of cognitive support for mapping tasks
- Develop design principles based on this model
- Develop and evaluate techniques to support these design principles in mapping tools

Visualizing Clinical Trials

- Developing new ways of visualizing clinical trials data
- Many different classes of users:
  - Clinicians, trialists, researchers
- Current approach for accessing trials is through pubmed or clinicaltrials.gov or RCT Presenter (Trial Bank)

Current Tools...

- [Image of RCT Presenter interface with trial data]
Visualizing Clinical Trials

- Developing new ways of visualizing clinical trials data
- Many different classes of users:
  - Clinicians, trialists, researchers
- Current approach for accessing trials is through pubmed or clinicaltrials.gov or RCT Presenter (Trial Bank)
- Currently striving for standard reporting procedures (ontology based)
- Visualization prototypes being developed for detailed and specific use cases

Limitations

- Use cases are highly specialized – trialist, researcher
- Requires a lot of information (which is currently not readily available)
- So what can we do with data that already preexists in ClinicalTrials.gov or PubMed?

Searching ClinicalTrials.gov

- How many trials are there related to the concept of "liver cancer"?
  - Cancer liver, 670
  - Carcinoma of Liver Cells, 344
  - Carcinoma, hepatocellular, 257
  - HCC, 411
  - Hepatocellular Carcinoma, 400
  - Hepatoma, 396
  - Liver cancer, 683
  - Liver Cell Carcinoma, 432
  - Liver metastases, 333
  - Metastases liver, 329
  - Primary Carcinoma of Liver Cells, 320

CTSearch

- Objective:
  - To improve the experience of seeking and comparing clinical trials
- Approach:
  - Use multiple tag clouds to support dynamic filtering and searching through clinical trials

Tag Clouds for summarizing Web Search Results
Tag Cloud visualization: CTSearch

- Browsing a larger set of trials through ClinicalTrials.gov
- Scenario
  - A trial designer wants to see all the conditions and interventions for "liver cancer". She wants to select one condition and see all the associated interventions (or vice versa).

Integration in BioPortal

TagSync approach (mock up ~ future work)

Summary

- Overview of NCBO
- Challenges – disparate ontologies, evolving, diversity of users/domains/tasks
  - Exploring visualization approaches in Jambalaya for BioPortal mashup services
  - Dealing with information overload and providing views that are useful (Diamond)
  - CogZ: Cognitive support for ontology alignment (and other potential applications of concept comparison)
  - Looking at specific areas for tool support – e.g. clinical trials, ICD 11 development

Future research directions and themes

- **Web based visualizations** for a diverse and evolving user base and task set
  - Customizable and flexible views, mashups
- **Cognitive support** rather than visualization
  - Ontologies, annotations and mappings
- **Collaborative support**
  - Humans and software agents

Input and suggestions would be greatly appreciated!

mstorey@uvic.ca

Computer Human Interaction & Software Engineering Lab
Department of Computer Science, University of Victoria
References and websites

- **NCBO**:
  - [http://bioontology.org](http://bioontology.org)

- **Jambalaya**:
  - [http://www.thechiselgroup.org/jambalaya](http://www.thechiselgroup.org/jambalaya)

- **Diamond**:
  - [http://thechiselgroup.org/diamond](http://thechiselgroup.org/diamond)

References and websites (2)

- **Visualization and Alignments**:

- **Clinical Trials**:
  - Trial Bank, RCT Presenter: [http://rctbank.ucsf.edu/](http://rctbank.ucsf.edu/)

Acknowledgements:
Chris Callendar, Tricia d'Entremont, Sean Falconer and Maleh Hernandez - The Chisel Group, UVic
Mark Musen, Principle Investigator NCBO – Stanford University

CiteULike

- **Features**:
  - The user queries one or more tags and receives a list of users and groups that are interested in those tags
  - Result: list of hyperlinks to articles, books or other sources of information
  - A tag cloud shows related tags for the user to use in further queries
  - Results are based on the 'union' rather than the 'intersection' of tags
  - Querying for 'Hepatocellular carcinoma' will bring all the resources tagged as 'Hepatocellular', 'carcinoma', or 'Hepatocellular carcinoma'. As a consequence, 'breast-carcinoma' will be part of the results.
  - The tag cloud can be filtered. For example, filtering on 'ca' displays: 'cancer', 'classification', 'communication', and 'education'. This only filters the cloud and not the list of results.