What needs to be represented in a biomedical ontology?

Olivier Bodenreider
National Library of Medicine
Bethesda, Maryland - USA

Anita Burgun
Medical School / Univ. Hospital
Rennes, France
Outline

◆ Granularity issues
◆ Types vs. instances
◆ Ontology vs. information model
Granularity issues
Granularity

◆ No theoretical limit
◆ No ideal granularity
◆ Arbitrary limitation
  ● Number of digits in code
  ● By design (purpose driven)
◆ Increased
  ● When several perspectives are combined (UMLS / individual vocabularies)
  ● When “hierarchy” is loosely defined
Granularity in UMLS

Distribution of depth max for 10,000 randomly selected UMLS concepts (2001)
Example  Addison’s disease

- Representation in several medical vocabularies
  - SNOMED International
  - Medical Subject Headings (MeSH)
  - Alcohol and Other Drugs Thesaurus
  - Read Codes (CTV3)
  - International Classification of Diseases

- Combined representation in the UMLS Metathesaurus
Diseases of the endocrine system

Diseases of the Adrenal Glands

Addison’s Disease
Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s Disease
Endocrine disorder

Adrenal disorder

Adrenal cortical disorder

Adrenal cortical hypofunction

Addison’s Disease
Endocrine disorder

Disorder of adrenal gland

Hypoadrenalism

Adrenal Hypofunction

Corticoadrenal insufficiency

Addison’s Disease
Adrenal Cortex Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Addison’s Disease

Endocrine Diseases

Adrenal Gland Diseases

organize concepts

SNOMED
MeSH
AOD
Read Codes

UMLS
Granularity and synonymy

- Granularity may be limited in order to represent only significant differences among concepts (linguistic synonymy vs. clinical synonymy)
- When granularity is limited by design, it may not be possible to represent some hierarchical relationships

Example: Acute infantile eczema in SNOMED
Lack of structure within a source

diseases of the skin and subcutaneous tissues

eczema

acute eczema

infantile eczema

acute infantile eczema
Granularity and redundancy

“Core” concepts

- Present in more than one vocabulary
- Essentially eliminates leaf nodes (structural equivalent)
- Also reduces the density of the graph
- Reduces the number of concepts dramatically (> 80%)
Types vs. Instances
Types and instances  Examples

◆ Types
  ● Liver *is a kind of* Organ
  ● Cirrhotic liver *is a kind of* Liver

◆ Instances
  ● Leipzig *is an instance of* City
  ● Barry Smith *is an instance of* Philosopher
  ● My liver *is an instance of* Liver
  ● This aspirin tablet *is an instance of* Clinical drug

Nothing can be *a kind of*  
- Leipzig  
- Barry Smith  
- My liver  
- This aspirin tablet
Types and instances  Biomedical domain

- **Types**
  - Terminologies
  - Ontologies

- **Instances**
  - Medical records
  - Patient databases

- **Classes**
  - Taxonomy
  - Classification

---

Class: Schizomycetes  
Order: Eubacteriales  
Family: Bacillaceae  
Genus: Clostridium  
Species: Clostridium botulinum
Types and instances  UMLS

◆ Two-level structure
  ● Semantic Network
    ■ 134 Semantic Types (STs)
    ■ Relationships among STs
  ● Metathesaurus
    ■ 800,000 concepts
    ■ Inter-concept relationships
  ● Link = categorization
    ■ Often isa
    ■ Rarely is an instance of
Types and instances  UMLS

◆ Essentially all types
  ● Semantic types
  ● Concepts

◆ Exceptions
  ● Named geographic areas
    Germany, Europe
  ● Named laws
    National Health Planning and Resources Development Act of 1974
  ● Named intellectual products
    Finnish translation of the Medical Subject Headings

◆ No explicit distinction
  between is a kind of and is an instance of
Types and instances  Identification

◆ Features of instances
  ● Structural: often leaf nodes
  ● Morphologic: often capitalized
  ● Syntactic: usually not preceded by a determiner

◆ Applications
  ● Named entity recognition

[E. Alfonseca, GWA 2002]
Types, instances and granularity

- Is there a gradient between type and instance?
- Is the hierarchy well-formed?
Types and instances

Summary

- Biomedical ontology
  - Essentially types
  - Some classes (taxonomy)
  - Hardly any instances
- Instanciation: medical records
- Relationships: must distinguish between
  - Is an kind of (type-type)
  - Is an instance of (instance-class, type-class)
Ontology vs. Information model
Aspirin revisited

[...] benzoic acids
hydroxybenzoic acids
salicylic acids
aspirin
aspirin oral tablet
aspirin oral tablet 500 mg

Clinical drug
- Has ingredient(s): aspirin
- Has dosage: 500 mg
- Has route: oral
- Has form: tablet

[RTM Drug Model, VA]
Other examples Medical procedures

◆ Medical procedure
  ● Action
  ● Anatomic site
  ● Instruments
  ● Approach

◆ Appendicectomy
  ● Remove
  ● Appendix
  ● Surgical instruments
  ● Open surgery

[MAOUSSC, GALEN-IN-USE, SNOMED-RT]
### Other examples  Lab/clinical results

<table>
<thead>
<tr>
<th>Lab test</th>
<th>Sodium measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Sodium</td>
</tr>
<tr>
<td>Property</td>
<td>Serum concentration</td>
</tr>
<tr>
<td>Time aspect</td>
<td>Point in time</td>
</tr>
<tr>
<td>System/Sample</td>
<td>Serum/Plasma</td>
</tr>
<tr>
<td>Scale</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Method</td>
<td>N/A</td>
</tr>
</tbody>
</table>

LOINC “terms”: SODIUM:SCNC:PT:SER/PLAS:QN

[LOINC, Regenstrief Institute]
Ontology vs. Information model

◆ Ontology
  ● What: Meaning
  ● How: hierarchies, frames, description logics
  ● How big: often very large (hundreds of thousands of concepts)
  ● Access: through browsers

◆ Information model
  ● What: Structure
  ● How: UML diagrams
  ● How big: the model is often limited in size
  ● Access: “readable” diagram
  ● Populated with concepts from the ontology

[A. Rector, MEDINFO 2001]
Although O. and I.M. may be equivalent

- **Appendicectomy**
  - Remove
  - Appendix
  - Surgical instruments
  - Open surgery

(simplified representation)

- **Appendicectomy**
  - uses surgical instruments

- **Surgical procedure**
  - has approach open surgical procedure

- **Open surgical procedure**
  - has location appendix

- **Medical procedure**
Contact information

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
olivier@nlm.nih.gov

Anita Burgun
Laboratoire d’Informatique Médicale
Université Rennes 1
Rennes - France
Anita.Burgun@univ-rennes1.fr