Objectives

◆ Learn about biomedical ontologies
  ● History
  ● Design principles, formalisms and tools
  ● What are they?
  ● What are they used for?

◆ Work with biomedical ontologies
  ● Search
  ● Analyze
  ● Extend
  ● Use for data integration
# Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
References Bio-ontology courses

- Barry Smith, U. Buffalo / NCBO
  - http://ontology.buffalo.edu/smith/Ontology_Course.html
- Stefan Schulz, U. Freiburg, Germany / KR-MED
  - 2008 tutorial
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Introduction to Biomedical Ontologies

Olivier Bodenreider

Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Outline

- Historical perspective
- Introduction to biomedical terminologies through an example
- Biomedical terms as names for biomedical classes
- Terminological relations as a surrogate for ontological relations
Historical perspective
Why biomedical terminologies?

- To support a theory of diseases
- To classify diseases
- To support epidemiology
- To index and retrieve information
- To serve as a reference
To support a theory of diseases

◆ Hippocrates
  - Dismisses superstition
  - Four humors
    - Blood
    - Phlegm
    - Yellow bile
    - Black bile

◆ Thomas Sydenham (1624-1689)
  - Medical observations on the history and cure of acute diseases (1676)
To classify diseases (and plants)

◆ Carolus Linnaeus (1707-1778)
  ● *Genera Plantarum* (1737)
  ● *Genera Morborum* (1763)

◆ François Boissier de La Croix
  a.k.a. F. B. de Sauvages (1706-1767)
  ● *Methodus Foliorum* (1751)
  ● *Nosologia Methodica* (1763/68)

◆ William Cullen (1710-1790)
  ● *Synopsis Nosologiae Methodicae* (1785)
From plants...
… to diseases

◆ Four categories (W. Cullen)
  - Fevers
  - Nervous disorders
  - Cachexias
  - Local diseases

“The distinction of the genera of diseases, the distinction of the species of each, and often even that of the varieties, I hold to be a necessary foundation of every plan of physic, whether dogmatical or empirical.”
– William Cullen, Edinburgh, 1785
Synopsis Nosologia Methodicae

(Cited by Chris Chute)
To support epidemiology

- **John Graunt (1620-1674)**
  - Analyzes the vital statistics of the citizens of London
- **William Farr (1807-1883)**
  - Medical statistician
  - Improves Cullen’s classification
  - Contributes to creating ICD
- **Jacques Berthillon (1851-1922)**
  - Chief of the statistical services (Paris)
  - Classification of causes of death (161 rubrics)
Limitations of existing classifications

“The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that it is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases: vague, inconvenient names have been employed, or complications have been registered instead of primary diseases. The nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences, and should be settled without delay.”

– William Farr

First annual report.
To index and retrieve information

◆ Biomedical literature
  ● MEDLINE (15M citations from 4600 journals)
  ● Manually indexed
  ● Medical Subject Headings (MeSH)

◆ Genome
  ● Model organism databases (Fly, Mouse, Yeast, …)
  ● Manually / semi-automatically curated
  ● Gene Ontology
Black bile and psychomotor retardation: shades of melancholia in Dante's Inferno.

Widmer DA.

Memorial Sloan-Kettering Cancer Center, New York, NY 10017, USA. widmerd@mskcc.org

The history of melancholy depression is rich with images of movement retardation and mental dysfunction. The recent restoration of psychomotor symptoms to the diagnostic terminology of affective disorder is not novel to the students of medieval melancholia. The move back to the biology of this psychomotor dysfunction with the technical advances in brain imaging in recent years only echoes centuries-old writings on the centrality of movement changes in the depressive condition. The Inferno, the first cantica of Dante Alighieri's Commedia, has a wonderful abundance of allusions to the importance of psychomotor symptoms in describing the depressed individual. Slowed steps, garbled speech, frozen tears, these and many other images keep the physical manifestations of psychomotor suffering in the forefront of the reader's mind. Considering Medieval and Renaissance writings on melancholy suffering, it is fitting that Dante shows a bodily illness reflected in the hellish torments visited on the damned. From the souls of the sullen to those of the violent, the panorama of psychomotor symptoms plays a prominent role in the poem as well as in the medical and literary prose of succeeding centuries.

MeSH Terms:
- Depressive Disorder/history*
- History of Medicine, Medieval
- Human
- Italy
- Literature, Medieval/history*
- Medicine in Literature*
- Poetry/history*
- Psychomotor Disorders/history*
Mouse Genome Database and GO

Entrez Gene

1: Nf2  neurofibromatosis 2  [Mus musculus]
GeneID: 18016  Locus tag: MGL97307

General gene information

GeneOntology
  Provided by MGI
  Function
    cytoskeletal protein binding
    protein binding
    structural molecule activity
  Process
    intercellular junction assembly and/or maintenance
    negative regulation of cell cycle
    negative regulation of protein kinase activity
    regulation of cell proliferation
  Component
    adherens junction
    cytoplasm
    cytoskeleton
    membrane
To serve as a reference

◆ Reference terminology/ontology
  • Universally needed
  • Developed independently of any purposes
  • Reusable by many applications

◆ Examples
  • VA National Drug File (NDF)
  • Foundational Model of Anatomy (FMA)
  • SNOMED CT
Anatomy in Biomedicine

- Physiology
- Clinical medicine
- Biomedical literature
- Biomedical research
Administrative terminologies

- Coding patient records
  - International Classification of Primary Care (ICPC)
  - SNOMED
  - Read Codes

- Reporting claims to health insurance companies
  - International Classification of Diseases (ICD-9 CM)
  - Healthcare Common Procedure Coding System (HCPCS)
History of Medical Ontologies

Synopsis
Nosologiae
Methodicae

ICD

1603
1700
1785
1855
1900
1975

OPCS
SNOP
CPT
EmTree
MeSH
ICD9

FMA
GALEN
DM&D

OPCS3
OPCS4
OPCS4.3

UMLS

READ

SNOMED
International
SNOMED-RT
SNOMED-CT

1975
1985
1995
2005

Lister Hill National Center for Biomedical Communications

[Bodenreider, BIB 2006] (courtesy of J. Rogers)
Biomedical ontology in PubMed

Number of articles in PubMed/MEDLINE on Ontology vs. Controlled vocabulary

- Ontology or ontologies
- Both
- Controlled vocabulary [excluding DSM]

(*) As of 2008/02/20
(Partial coverage for 2007, due to a slight lag in the indexing process)

[Bodenreider, YBMI 2008]
Biomedical ontologies in PubMed

Proportion of citations in PubMed/MEDLINE by ontology

- Gene Ontology
- UMLS
- SNOMED
- MeSH
- LOINC

[Bodenreider, YBMI 2008]
Introduction to biomedical terminologies through an example
Guy’s Hospital, London
Thomas Addison (1795-1860)
Addison’s disease

◆ Addison's disease is a rare endocrine disorder
◆ Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
◆ For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism
Adrenal insufficiency  Clinical variants

- Primary / Secondary
  - Primary: lesion of the adrenal glands themselves
  - Secondary: inadequate secretion of ACTH by the pituitary gland
- Acute / Chronic
- Isolated / Polyendocrine deficiency syndrome
Addison’s disease: Symptoms

- Fatigue
- Weakness
- Low blood pressure
- Pigmentation of the skin (exposed and non-exposed parts of the body)
- ...
AD in medical vocabularies

◆ Synonyms: different terms
  • Addisonian syndrome
  • Bronzed disease
  • Addison melanoderma
  • Asthenia pigmentosa
  • Primary adrenal deficiency
  • Primary adrenal insufficiency
  • Primary adrenocortical insufficiency
  • Chronic adrenocortical insufficiency

◆ Contexts: different hierarchies

  eponym
  symptoms
  clinical
  variants
CHAPTER 4
Endocrine, nutritional and metabolic diseases (E00-E90)

Disorders of other endocrine glands (E20-E35)

E27 Other disorders of adrenal gland
  E27.0 Other adrenocortical overactivity
              Overproduction of ACTH, not associated with Cushing's disease
              Premature adrenarche
              Excludes1: Cushing's syndrome (E24.-)
  E27.1 Primary adrenocortical insufficiency
             Addison's disease
             Adrenocortical insufficiency NOS
             Autoimmune adrenalitis
             Excludes1: Addison only phenotype adrenoleukodystrophy (E71.428)
                     amyloidosis (E85)
                     tuberculous Addison's disease (A18.7)
                     Waterhouse-Friderichsen syndrome (A39.1)
  E27.2 Addisonian crisis
             Adrenal crisis
             Adrenocortical crisis
  E27.3 Drug-induced adrenocortical insufficiency
             Code first (T36-T50) to identify drug
  E27.4 Other and unspecified adrenocortical insufficiency
Medical Subject Headings

MeSH Tree Structures

Endocrine Diseases [C:9]
  Adrenal Gland Diseases [C:19.053]
    Adrenal Gland Hypofunction [C:19.053.264]
      Addison's Disease [C:19.053.264.263]
    Adrenoleukodystrophy [C:19.053.264.270]
    Hypoaldosteronism [C:19.053.264.480]

Immunologic Diseases [C:20]
  Autoimmune Diseases [C:20.111]
    Addison's Disease [C:20.111.163]
    Anemia, Hemolytic, Autoimmune [C:20.111.175]
    Anti-Glomerular Basement Membrane Disease [C:20.111.190]
    Antiphospholipid Syndrome [C:20.111.197]
    Arthritis, Rheumatoid [C:20.111.199]
    Autoimmune Diseases of the Nervous System [C:20.111.258]
Addison's disease - Definition

Concept Status: Current

Descriptions
- Addison's disease (disorder)
- Addison's disease
- Enfermedad de Addison
- Enfermedad de Addison (trastorno)

Definition: Primitive
- is a
  - Adrenal cortical hypofunction
  - Finding site
  - Adrenal cortex structure

Qualifiers
- Severity
  - Severe
  - Severe
  - Episodic
  - Episodic
  - Clinical course
  - Courses

Codes
- Original SnomedId: DB-70620
- Read Code (Ctv3Id): C1541

Hierarchy
Subtype hierarchy

386594007 adrenal cortical hypofunction
383732003 Addison's disease
38760008 Addison's disease with adrenoleucodystrophy
78715008 Addison's disease due to autoimmunity
186270000 tuberculous Addison's disease
11244009 polyglandular autoimmune syndrome, type 1
Biomedical terms as names for biomedical classes
Terms reflecting valid classes

- Pulmonary anthrax
- BRCA1 protein
- Coronary artery
- Coronary artery bypass
- ...

- Non-insulin dependent diabetes mellitus
- Non-Hodgkin lymphoma
- Non-steroidal anti-inflammatory drugs
- Non-opioid analgesics
- Non-invasive medical procedure
Issues

- Multiple terms for a class
- Multiple classes for a term
- Presence of non-ontological features in terms
- Composite terms
Multiple terms for a class

◆ Synonymy
  - Left coronary artery
  - LCA
  - Arteria coronaria sinistra
  - Addison’s disease
  - Primary adrenocortical insufficiency

◆ “Clinical synonymy” (vs. identity)
  - Abdominal swelling
  - Swollen abdomen
  - Posttransfusion hepatitis
  - Posttransfusion viral hepatitis
  - Addison’s disease
  - Primary adrenocortical insufficiency
  - vs. Waterhouse-Friderichsen Syndrome
Multiple classes for a term

- Polysemy
  - Cold
    - Cold
    - Cold temperature
  - Common cold
  - COLD
    - Chronic Obstructive Airway Disease

- Truncated terms
  - Calcium
    - Ca++
    - Coagulation factor IV
  - Calcium measurement
Non-ontological features in terms of ontological features

- Epistemological features
  - Gallbladder calculus without mention of cholecystitis
  - Diarrhea of presumed infectious origin
  - Replacement of unspecified heart valve
  - …
Ontology vs. Epistemology

- **Ontology**
  - Invariants in reality
    - Classes (universals)
    - Relations between them
  - Theory of reality

- **Epistemology**
  - Knowledge about such entities
  - Perception of reality

Bone metastasis
- diagnosed by CT scan
- diagnosed by Tc99m bone scintiscan
Composite terms

◆ Sentence-like terms
   - Several classes and their relations
   - May contain epistemological features

   ▪ Tuberculosis of adrenal glands, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture
More composite terms

- Nontraffic accident involving being accidentally pushed from motor vehicle, except off-road motor vehicle, while in motion, not on public highway, driver of motor vehicle injured

- Determine whether the elder patient and caretaker have a functional social support network to assist the patient in performing activities of daily living and in obtaining health care, transportation, therapy, medications, community resource information, financial advice, and assistance with personal problems

- Telephone call by a physician to patient or for consultation or medical management or for coordinating medical management with other health care professionals (eg, nurses, therapists, social workers, nutritionists, physicians, pharmacists); complex or lengthy (eg, lengthy counseling session with anxious or distraught patient, detailed or prolonged discussion with family members regarding seriously ill patient, lengthy communication necessary to coordinate complex services of several different health professionals working on different
Terminological relations as a surrogate for ontological relations
Issues

- Lack of explicit classificatory principle
- Underspecification of the relations
- Thesaurus relations
- Limited depth in hierarchies “by design”
Explicit classificatory principle

Foundational Model of Anatomy

Anatomical entity

- Spatial dimension
- Mass

Material physical anatomical entity

- Inherent 3D shape
- Body substance

Non-material physical anatomical entity

Anatomical structure

Body substance

- Anat. space
- Anat. surface
- Anat. line
- Anat. point

3D 2D 1D 0D

Lister Hill National Center for Biomedical Communications
No explicit classificatory principle
1. Certain infectious and parasitic diseases
2. Neoplasms
3. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4. Endocrine, nutritional, and metabolic diseases
5. Mental and behavioral disorders
6. Diseases of nervous system
7. Diseases of the eye and adnexa
8. Diseases of the ear and mastoid process
9. Diseases of circulatory system
10. Diseases of respiratory system
11. Diseases of digestive system
12. Diseases of the skin and subcutaneous tissue
13. Diseases of the musculoskeletal system and connective tissue
14. Diseases of the genitourinary system
15. Pregnancy, childbirth, and the puerperium
16. Certain conditions originating in the newborn (perinatal) period
17. Congenital malformations, deformations and chromosomal abnormalities
18. Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19. Injury, poisoning and certain other consequences of external causes
20. External causes of morbidity
21. Factors influencing health status and contact with health service
• Attribute
• Body structure
• Clinical finding
• Context-dependent categories
• Environments and geographical locations
• Events
• Observable entity
• Organism
• Pharmaceutical / biologic product
• Physical force
• Physical object
• Procedure
• Qualifier value
• Social context
• Special concept
• Specimen
• Staging and scales
• Substance
Fully specified relations

Viral meningitis in SNOMED CT

- Fully defined by ...
  - Is a
    - D viral infections of the central nervous system
    - D infective meningitis
  - Causative agent: D virus
  - Group
    - Associated morphology: D inflammation
    - Finding site: D meninges structure
Underspecification of the relations

Diseases

CNS diseases

Virus diseases

CNS infections

Meningitis

CNS viral diseases

Viral meningitis

parent

isa ?

child
Thesaurus relations

◆ Addison’s disease
  ● Due to auto-immunity in 80% of the cases
  ● Other causes include tuberculosis

Relations used to create hierarchical structures vs. hierarchical relations
Endocrine Diseases [C19]

Adrenal Gland Diseases [C19.053]
  Adrenal Gland Hypofunction [C19.053.264]
   ➤ Addison's Disease [C19.053.264.263]
  Adrenoleukodystrophy [C19.053.264.270]
  Hypoaldosteronism [C19.053.264.480]

Immunologic Diseases [C20]

Autoimmune Diseases [C20.111]
  ➤ Addison's Disease [C20.111.163]
  Anemia, Hemolytic, Autoimmune [C20.111.175]
  Anti-Glomerular Basement Membrane Disease [C20.111.190]
  Antiphospholipid Syndrome [C20.111.197]
  Arthritis, Rheumatoid [C20.111.199] +

Hierarchy

adrenal cortical hypofunction
  ➤ Addison's disease
    ➤ Addison's disease due to autoimmunity
    ➤ Addison's disease with adrenoleukodystrophy
    ➤ polyglandular autoimmune immune syndrome, type 1
    ➤ tuberculous Addison's disease
Accidents in MeSH

Environment and Public Health [G03]
Public Health [G03.850]

Accidents [G03.850.110]

- Accident Prevention [G03.850.110.060] +
- Accidental Falls [G03.850.110.085]
- Accidents, Aviation [G03.850.110.185]
- Accidents, Home [G03.850.110.205]
- Accidents, Occupational [G03.850.110.250] +
- Accidents, Radiation [G03.850.110.285]
- Accidents, Traffic [G03.850.110.320]
- Drowning [G03.850.110.500] +
Limited depth in hierarchies “by design”

◆ Term identifier (code) used to record the position in the hierarchy
  - Limited number of digits available
  - May hide part of the structure
◆ Terminologies: ICD, SNOMED, …

E84 Cystic fibrosis
   Includes: mucoviscidosis
   E84.0 Cystic fibrosis with pulmonary manifestations
      Use additional code to identify any infectious organism present, such as: Pseudomonas (B96.5)
   E84.1 Meconium ileus in cystic fibrosis
      Excludes1: meconium ileus not due to Cystic fibrosis (P75)
   E84.2 Cystic fibrosis with gastrointestinal manifestations
      Excludes2: meconium ileus in cystic fibrosis (E84.1)
   E84.8 Cystic fibrosis with other manifestations
Cystic fibrosis in ICD

E84  Cystic fibrosis
    Includes: mucoviscidosis
E84.0 Cystic fibrosis with pulmonary manifestations
    Use additional code to identify any infectious organism present, such as:
      Pseudomonas (B96.5)
E84.1 Meconium ileus in cystic fibrosis
    Excludes1: meconium ileus not due to Cystic fibrosis (P75)
E84.2 Cystic fibrosis with gastrointestinal manifestations
    Excludes2: meconium ileus in cystic fibrosis (E84.1)
E84.8 Cystic fibrosis with other manifestations

Lister Hill National Center for Biomedical Communications
Conclusions
Conclusions 😞

◆ Biomedical terms
  ● reflect some aspects of biomedical reality
    ■ Although the primary concern of terminology is naming, not reflecting reality
  ● often convey additional features (e.g., epistemology)

◆ Biomedical terminology tends to offset part of the complexity
  ● but often reflects utility
Conclusions 😊

◆ Biomedical terminologies can help populate biomedical ontologies

◆ Resources needed
  ● Linguistic analysis of terms
  ● Statistical analysis of terms in a corpus / annotation database (dependence relations)
  ● Manual curation
## Agenda

<table>
<thead>
<tr>
<th></th>
<th>Monday, June 9</th>
<th>Tuesday, June 10</th>
<th>Wednesday, June 11</th>
</tr>
</thead>
</table>
| **Introduction to Biomedical Ontologies** | Design Principles, Formalisms and Tools for Biomedical Ontologies | Biomedical Ontologies
  - Content and structure
  - Function |
| Interfaces to Biomedical Ontologies | Searching and Analyzing Biomedical Concepts | Contrasting Biomedical Ontologies |
| Critical Analysis of Biomedical Ontologies | Extending Biomedical Ontologies | Using Biomedical Ontologies
  for Data Integration |
Design Principles, Formalisms and Tools for Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Definitions
  ● Ontologies vs. other artifacts
  ● Kinds of ontologies
◆ Some principles of formal ontology
  ● Top-level categories
  ● Categories of relationships
◆ Formalisms and tools
Definitions
Introduction

Ogden-Richards

Symbol

Object
Definitions

◆ The *What* question
  - Objects in the world
    - With their properties
    - With their relations to other objects
  - Also: events, processes, and states

◆ Explicit specification of a conceptualization
  - Support software applications

◆ Domain ontology reflects
  - Underlying reality
  - Theory of the domain
Examples of use

◆ Natural language processing
◆ Access to heterogeneous sources of information (e.g., Semantic Web)
◆ Systems engineering

◆ Interoperability
Ontology vs. other artifacts

◆ Ontology
  ● Defining types of things and their relations
◆ Terminology
  ● Naming things in a domain
◆ Thesaurus
  ● Organizing things for a given purpose
◆ Classification
  ● Placing things into (arbitrary) classes
◆ Knowledge bases
  ● Assertional knowledge

[Smith, KR-MED 2006]
[Chute, JAMIA 2000]
(Controlled) Terminology

◆ Objective: naming things
◆ Shared understanding
  ● Agreement on what terms to use
  ● Utility-driven (arbitrary)

Telephone call by a physician to patient or for consultation or medical management or for coordinating medical management with other health care professionals (eg, nurses, therapists, social workers, nutritionists, physicians, pharmacists); complex or lengthy (eg, lengthy counseling session with anxious or distraught patient, detailed or prolonged discussion with family members regarding seriously ill patient, lengthy communication necessary to coordinate complex services of several different health professionals working on different
Thesaurus

◆ Objective: organize things for a purpose
  ● e.g., information retrieval
    ■ Organization by relatedness
◆ Example: Medical Subject Headings (MeSH)
  ● Indexing/retrieval of biomedical articles
◆ Relations used in hierarchies
  vs. hierarchical relations
Thesaurus vs. ontology

Autoimmune Diseases

is generally a

Addison’s disease

Tuberculous Addison’s disease

Addison’s disease due to autoimmunity
Classification

◆ Objective: placing things into classes

◆ Characteristics
  • Single inheritance (tree)
  • Idiosyncratic inclusion/exclusion criteria

**Insulin-dependent diabetes mellitus**

[See before E10 for subdivisions.]

**Includes:**
- brittle
- juvenile-onset
- ketosis-prone
- type I

**Excludes:**
- malnutrition-related (E12-)
- neonatal (P70.2)
- pregnancy, childbirth and the puerperium (Q24-)
- glycosuria:
  - NOS (R81)
  - renal (E74.8)
- impaired glucose tolerance (R73.0)
- postsurgical hypoinsulinaemia (E89.1)
Classification

◆ Characteristics (continued)
  - Everything must be classified, including
    - When there is no specific slot (NEC)
    - When there is insufficient information (NOS)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E84</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td></td>
<td><em>Includes: mucoviscidosis</em></td>
</tr>
<tr>
<td>E84.0</td>
<td>Cystic fibrosis with pulmonary manifestations</td>
</tr>
<tr>
<td>E84.1</td>
<td>Cystic fibrosis with intestinal manifestations</td>
</tr>
<tr>
<td></td>
<td>Meconium ileus+ (P75*)</td>
</tr>
<tr>
<td></td>
<td><em>Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)</em></td>
</tr>
<tr>
<td>E84.8</td>
<td>Cystic fibrosis with other manifestations</td>
</tr>
<tr>
<td>E84.9</td>
<td>Cystic fibrosis, unspecified</td>
</tr>
</tbody>
</table>
Knowledge Bases

◆ Objective: represent knowledge needed for a given application
◆ Example: drug knowledge bases
◆ Assertional knowledge
  • Vs. definitional knowledge in ontologies
  • Often probabilistic
◆ Examples of assertions
  • Indications of a drug
  • Signs and symptoms of a disease
Fuzzy borders

- Some ontologies also collect names
  - FMA
- Some terminologies also provide formal definitions
  - SNOMED CT
- Some terminologies/ontologies include both definitional and assertional knowledge
  - SNOMED CT
Types of resources

- **Lexical resources**
  - Collections of lexical items
  - Additional information
    - Part of speech
    - Spelling variants
  - Useful for entity recognition
  - UMLS SPECIALIST Lexicon, WordNet

- **Ontological resources**
  - Collections of
    - kinds of entities
      (substances, qualities, processes)
    - relations among them
  - Useful for relation extraction
  - UMLS Semantic Network, BioTop

- **Terminological resources**
  - Collections lexical items + identifiers
    - Useful for entity resolution
    - UMLS Metathesaurus
The Knowledge Semantics Continuum

Medication Lists, DDI Lists

Thesauri: BT/NT, Parent/Child, Informal Is-A

Catalog

Terms/glossary

MeSH, Gene Ontology, UMLS Meta

DB Schema

KEGG

TAMBIS

BioPAX

Formal is-a Frames (Properties)

RDF(S)

Formal instances

Disjointness, Inverse

Ontylog

OWL

IEEE SUO

General Logical constraints

Snomed

EcoCyc

Simple Terminologies

Expressive Ontologies

Ontology Dimensions based on McGuinness and Finin
Kinds of ontologies

- Upper Level Ontology
- General Ontology
- Domain Ontology

Application ontologies
Ontology-related issues

- Creation
- Merging
- Alignment
- Validation
Formal Ontological Principles
Formal ontological distinctions

- Universal vs. individual
- Continuant vs. occurrent
- Independent vs. dependent
Universal vs. Individual

- Universal = category
- Synonyms
  - Kind, Type, (Class)
- Examples
  - eyeball
  - blood pressure
  - conference

- Individual = instance
- Synonyms
  - Particular, Token
- Examples
  - my right eyeball
  - my blood pressure (132/79)
  - AMIA Annual Symposium 2003

instantiation
Continuant vs. Occurrent

- Continuant = \textit{Continues to exist through time}
- **Synonyms**
  - Substance
- **Examples**
  - tennis racquet
  - mitochondrion
  - insulin production

- Occurrent = \textit{Unfolds through time}
- **Synonyms**
  - Process
- **Examples**
  - tennis tournament
  - metabolism
  - producing insulin
Independent vs. Dependent

- Independent = *Can exist without support from other entities*
  - Examples
    - virus
    - molecule
    - plant

- Dependent = *Require support from other entities for its existence*
  - Examples
    - viral infection
    - DNA binding
    - food
Formal ontology  Upper level

- **Universals** (classes)
  - Continuant
    - Independent continuant
    - Dependent continuant
  - Occurent

- **Particulars** (instances)

Lister Hill National Center for Biomedical Communications
Formal ontological distinctions

- Basic distinctions in many top-level ontologies
  - Generic: BFO, DOLCE
  - Biomedical: BioTop, UMLS Semantic Network

- Condition the relations between various types of entities
  - Relations
    - Between instances (e.g., part_of [at time])
    - Between classes (e.g., isa, part_of [atemporonal])
    - Between one instance and one class (instance_of)

[Smith, Genome Biology 2005]
Formal ontology in practice

- Provides foundational classes and relations
  - Upper level ontologies
  - Relation ontology
- Provides a framework for analyzing entities and relations
Examples
General ontologies

- **OpenCyc**
  - General ontology
  - Cycorp, Inc (D. Lenat & al.)
  - Over 1M hand-coded assertions
  - [http://www.opencyc.org](http://www.opencyc.org)

- **WordNet**
  - Electronic lexical database
  - Princeton University (G. Miller & al.)
  - Over 100,000 synsets
  - [http://wordnet.princeton.edu/](http://wordnet.princeton.edu/)
Top level in OpenCyc
Top level in WordNet

- Abstraction
- Activity
- Entity
- Event
- Group
- Location
- Natural phenomenon
- Possession
- Psychological feature
- Shape
- State
GALEN

- Generalised Architecture for Languages, Encyclopaedias, and Nomenclatures in Medicine
- European Union project (A. Rector & al.)
- Over 25,000 concepts (primitives)
- http://www.opengalen.org
Top level in GALEN

![Diagram showing the hierarchical structure of GALEN concepts]

- **DomainCategory**
  - **Phenomenon**
    - **Generalised Structure**
    - **Generalised Process**
    - **Generalised Substance**
    - **SubstanceOr PhysicalStructure**
  - **Arbitrarily Conjuncted Phenomenon**
  - **ModifierConcept**
    - **Collection**
    - **Aspect**
      - **Feature**
      - **StateOr Quantity**
      - **Status**
      - **Selector**
    - **GeneralLevel OfSpecification**
    - **Unit**
    - **Role**
    - **Modality**

Lister Hill National Center for Biomedical Communications
UMLS Semantic Network

- Definitional knowledge in the biomedical domain
- NLM (A. McCray & al.)
- Content
  - 135 semantic types
  - 54 types of relationship
  - 6700 semantic relations
Top level in the Semantic Network
Differences between ontologies

Examples
Granularity, plesionymy

UMLS

- Epilepsy, Generalized
- Seizure Disorder, Generalized
- [...]
Differing categorization

UMLS

Natural Phenomenon or Process

Biologic Function

Pathologic Function

Disease or Syndrome

Dental Caries
Dental cavity, NOS
Tooth caries
Dental Decay
[...]

WordNet

phenomenon

process

natural process

decay

cavity caries
dental caries
tooth decay

Lister Hill National Center for Biomedical Communications
Formalisms and Tools
Ontology and Formalism

- Frames
- Description logics
  - OWL DL
- First-order logic

- OBO Format
  - Conversion to OWL DL
Tools for ontology developers

◆ Protégé
  ● Publicly available
  ● Frames and DL
  ● Classifier
  ● Supports many file formats (import/export)
  ● Large community of users
  ● Well supported
  ● http://protege.stanford.edu/

◆ OBO-Edit
  ● Specific to the biomedical/OBO community
  ● Simpler than Protégé (“tool for biologists”)
  ● http://oboedit.org/

Lister Hill National Center for Biomedical Communications
<table>
<thead>
<tr>
<th>Day</th>
<th>Agenda Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, June 9</td>
<td>Introduction to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>- Content and structure</td>
</tr>
<tr>
<td></td>
<td>- Function</td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Searching and Analyzing Biomedical Concepts</td>
</tr>
<tr>
<td></td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Extending Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>

Lister Hill National Center for Biomedical Communications
“High-Impact” Biomedical Ontologies

A Structural Perspective

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Structural perspective
  - What are they (vs. what are they for)?

◆ “High-impact” biomedical ontologies
  - International Classification of Diseases (ICD)
  - Logical Observation Identifiers, Names and Codes (LOINC)
  - SNOMED Clinical Terms
  - Foundational Model of Anatomy
  - Gene Ontology
  - RxNorm
  - Medical Subject Headings (MeSH)
  - NCI Thesaurus
  - Unified Medical Language System (UMLS)

[J. Cimino, YBMI 2006]
International Classification of Diseases
ICD Characteristics (1)

- Current version: ICD-10
- Type: Classification
- Domain: Disorders
- Developer: World Health Organization (WHO)
- Funding: WHO
- Availability
  - Publicly available: No
  - Repositories: UMLS  [ICD9-CM in NCBO BioPortal]
- URL: http://www.who.int/classifications/icd/en/
ICD Characteristics (2)

◆ Number of
  ● Concepts: 12,318
  ● Terms: 1 per concept (tabular)

◆ Major organizing principles:
  ● Tree (single inheritance hierarchy)
  ● No explicit classification criteria
    ■ Idiosyncratic inclusion/exclusion mechanism
  ● .8 slots for Not elsewhere classified (NEC)
  ● .9 slots for Not otherwise specified (NOS)

◆ Formalism: Proprietary format
# ICD Top level

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Blocks</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A00-B99</td>
<td>Certain infectious and parasitic diseases</td>
</tr>
<tr>
<td>II</td>
<td>C00-D48</td>
<td>Neoplasms</td>
</tr>
<tr>
<td>III</td>
<td>D50-D89</td>
<td>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism</td>
</tr>
<tr>
<td>IV</td>
<td>E00-E90</td>
<td>Endocrine, nutritional and metabolic diseases</td>
</tr>
<tr>
<td>V</td>
<td>F00-F99</td>
<td>Mental and behavioural disorders</td>
</tr>
<tr>
<td>VI</td>
<td>G00-G99</td>
<td>Diseases of the nervous system</td>
</tr>
<tr>
<td>VII</td>
<td>H00-H59</td>
<td>Diseases of the eye and adnexa</td>
</tr>
<tr>
<td>VIII</td>
<td>H60-H95</td>
<td>Diseases of the ear and mastoid process</td>
</tr>
<tr>
<td>IX</td>
<td>I00-I99</td>
<td>Diseases of the circulatory system</td>
</tr>
<tr>
<td>X</td>
<td>J00-J99</td>
<td>Diseases of the respiratory system</td>
</tr>
<tr>
<td>XI</td>
<td>K00-K93</td>
<td>Diseases of the digestive system</td>
</tr>
<tr>
<td>XII</td>
<td>L00-L99</td>
<td>Diseases of the skin and subcutaneous tissue</td>
</tr>
<tr>
<td>XIII</td>
<td>M00-M99</td>
<td>Diseases of the musculoskeletal system and connective tissue</td>
</tr>
<tr>
<td>XIV</td>
<td>N00-N99</td>
<td>Diseases of the genitourinary system</td>
</tr>
<tr>
<td>XV</td>
<td>O00-O99</td>
<td>Pregnancy, childbirth and the puerperium</td>
</tr>
<tr>
<td>XVI</td>
<td>P00-P96</td>
<td>Certain conditions originating in the perinatal period</td>
</tr>
<tr>
<td>XVII</td>
<td>Q00-Q99</td>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
</tr>
<tr>
<td>XVIII</td>
<td>R00-R99</td>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified</td>
</tr>
<tr>
<td>XIX</td>
<td>S00-T98</td>
<td>Injury, poisoning and certain other consequences of external causes</td>
</tr>
<tr>
<td>XX</td>
<td>V01-Y98</td>
<td>External causes of morbidity and mortality</td>
</tr>
<tr>
<td>XXI</td>
<td>Z00-Z99</td>
<td>Factors influencing health status and contact with health services</td>
</tr>
<tr>
<td>XXII</td>
<td>U00-U99</td>
<td>Codes for special purposes</td>
</tr>
</tbody>
</table>
ICD Example

- Idiosyncratic inclusion/exclusion criteria

**E10**

*Insulin-dependent diabetes mellitus*

[See before E10 for subdivisions.]

**Includes:**
- diabetes (mellitus):
  - brittle
  - juvenile-onset
  - ketosis-prone
  - type I

**Excludes:**
- diabetes mellitus (in):
  - malnutrition-related (E12.-)
  - neonatal (P70.2)
  - pregnancy, childbirth and the puerperium (O24.-)
  - glycosuria:
    - NOS (R81.)
    - renal (E74.8)
  - impaired glucose tolerance (R73.0)
  - postsurgical hypoinsulinaemia (E89.1)
ICD Example

- Not elsewhere classified (NEC)
- Not otherwise specified (NOS)

E84  Cystic fibrosis
  Includes: mucoviscidosis
E84.0  Cystic fibrosis with pulmonary manifestations
E84.1  Cystic fibrosis with intestinal manifestations
  Meconium ileus+ ( P75*)
  Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present ( P76.0)
E84.8  Cystic fibrosis with other manifestations
E84.9  Cystic fibrosis, unspecified
Logical Observation Identifiers, Names and Codes (LOINC)
LOINC Characteristics (1)

- Current version: 2.22 (Dec. 2007)
- Type: Controlled terminology*
- Domain: Laboratory and clinical observations
- Developer: Regenstrief Institute
- Funding: NLM
- Availability
  - Publicly available: Yes
  - Repositories: UMLS
- URL: www.regenstrief.org/loinc/loinc.htm
LOINC Characteristics (2)

- Number of
  - Concepts: 50k active codes (2.18)
  - Terms: n/a*

- Major organizing principles:
  - No hierarchical structure among the main codes
  - 6 axes
    - Component (analyte [+ challenge] [+ adjustments])
    - Property
    - Timing
    - System
    - Scale
    - [Method]

- Formalism: “DL-like”
LOINC Example

- \textit{Sodium}:SCnc:-Pt:Ser/Plas:Qn
  [the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

<table>
<thead>
<tr>
<th>Axis</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Sodium</td>
</tr>
<tr>
<td>Property</td>
<td>SCnc – Substance Concentration (per volume)</td>
</tr>
<tr>
<td>Timing</td>
<td>Pt – Point in time (Random)</td>
</tr>
<tr>
<td>System</td>
<td>Ser/Plas – Serum or Plasma</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn – Quantitative</td>
</tr>
<tr>
<td>Method</td>
<td>--</td>
</tr>
</tbody>
</table>
SNOMED Clinical Terms
SNOMED CT  Characteristics (1)

- Current version: January 31, 2008 (2 annual releases)
- Type: Reference terminology / ontology
- Domain: Clinical medicine
- Developer: IHTSDO
- Funding: IHTSDO
- Availability
  - Publicly available: Yes* (in member countries)
  - Repositories: UMLS
- URL: http://www.ihtsdo.org/
SNOMED CT Characteristics (2)

- Number of
  - Concepts: 311,313 active concepts (Jan. 31, 2008)
  - Terms: 794,061 active “descriptions”

- Major organizing principles:
  - Utility for clinical medicine (e.g., assertional + definitional knowledge)
  - Model of meaning (incomplete)
  - Rich set of associative relationships
  - Small proportion of defined concepts (many primitives)

- Formalism: Description logics (KRSS)
SNOMED CT  Top level

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>Subtype hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>138875005 SNOMED CT Concept</td>
</tr>
<tr>
<td></td>
<td>362981000 qualifier value</td>
</tr>
<tr>
<td></td>
<td>106237007 linkage concept</td>
</tr>
<tr>
<td></td>
<td>370115009 special concept</td>
</tr>
<tr>
<td></td>
<td>48176007 social context</td>
</tr>
<tr>
<td></td>
<td>419891008 record artifact</td>
</tr>
<tr>
<td></td>
<td>363787002 observable entity</td>
</tr>
<tr>
<td></td>
<td>308916002 environment or geographical location</td>
</tr>
<tr>
<td></td>
<td>123038009 specimen</td>
</tr>
<tr>
<td></td>
<td>254291000 staging and scales</td>
</tr>
<tr>
<td></td>
<td>123037004 body structure</td>
</tr>
<tr>
<td></td>
<td>272379006 event</td>
</tr>
<tr>
<td></td>
<td>78621006 physical force</td>
</tr>
<tr>
<td></td>
<td>404684003 clinical finding</td>
</tr>
<tr>
<td></td>
<td>280787004 physical object</td>
</tr>
<tr>
<td></td>
<td>410607006 organism</td>
</tr>
<tr>
<td></td>
<td>71388002 procedure</td>
</tr>
<tr>
<td></td>
<td>373873005 pharmaceutical / biologic product</td>
</tr>
<tr>
<td></td>
<td>243796009 situation with explicit context</td>
</tr>
<tr>
<td></td>
<td>105590001 substance</td>
</tr>
</tbody>
</table>
**SNOMED CT Example**

**Hierarchy**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>27010001</td>
<td>partial excision of large intestine</td>
</tr>
<tr>
<td>8613002</td>
<td>operation on appendix</td>
</tr>
<tr>
<td>80146002</td>
<td>appendectomy</td>
</tr>
<tr>
<td>82730006</td>
<td>incidental appendectomy</td>
</tr>
<tr>
<td>49438003</td>
<td>appendectomy with drainage</td>
</tr>
<tr>
<td>174036004</td>
<td>emergency appendectomy</td>
</tr>
<tr>
<td>174045003</td>
<td>interval appendectomy</td>
</tr>
<tr>
<td>6025007</td>
<td>laparoscopic appendectomy</td>
</tr>
<tr>
<td>235313004</td>
<td>non-emergency appendectomy</td>
</tr>
<tr>
<td>235314005</td>
<td>inversion appendectomy</td>
</tr>
<tr>
<td>1299000</td>
<td>excision of appendiceal stump</td>
</tr>
</tbody>
</table>

**Definition**

*Fully defined by...*

- **Method**
  - excision - action
- **procedure site** - Direct
- **appendix structure**

**Qualifiers**

- access
  - surgical access values
- priority
  - priorities

**Codes**

- Original SnomedId: P1-57450
- Read Code (Ctv3Id): X20Wz
Foundational Model of Anatomy
FMA Characteristics (1)

- Current version: ? (no fixed release schedule)
- Type: Ontology
- Domain: Anatomy (anatomical structures)
- Developer: U. Washington, Department of Biological Structure
- Funding: NLM (grants and contract) and others
- Availability
  - Publicly available: Yes
  - Repositories: [UMLS] / OBO / NCBO BioPortal
- URL: http://fma.biostr.washington.edu/
FMA Characteristics (2)

- **Number of**
  - Concepts: ~72k
  - Terms: ~1.5 / concept

- **Major organizing principles:**
  - Explicit classificatory criteria
  - Distinct *isa* and *part_of* hierarchies
  - Additional spatial relations (e.g., adjacency)
  - Multiple levels of granularity (organism to sub-cellular)

- **Formalism: Frames (Protégé)**
  - Conversion to OWL Full and OWL DL available
FMA Top level

(Courtesy of C. Rosse)
FMA Example

(Courtesy of C. Rosse)
Gene Ontology
Gene Ontology  Characteristics (1)

- Current version: n/a (daily/monthly releases)
- Type: Controlled vocabulary
- Domain: Molecular biology
- Developer: GO Consortium
- Funding: NIH (grants)

**Availability**
- Publicly available: Yes
- Repositories: UMLS / OBO / NCBO BioPortal

**URL:** http://www.geneontology.org/
Gene Ontology  Characteristics (2)

◆ Number of
  ● Terms: 2.15 per concept

◆ Major organizing principles:
  ● 3 major hierarchies
    ● Molecular function
    ● Cellular component
    ● Biological process
  ● Relations (within hierarchies): \(isa, part\_of,\) regulates
  ● No relations between concepts across hierarchies

◆ Formalism: OBO format
Gene Ontology  Top level (MF)

- all : all [250418 gene products]
  - GO:0008150 : biological_process [166605 gene products]
  - GO:0005575 : cellular_component [169814 gene products]
  - GO:0003674 : molecular_function [168558 gene products]
    - GO:0016209 : antioxidant activity [566 gene products]
    - GO:0015457 : auxiliary transport protein activity [161 gene products]
    - GO:0005488 : binding [46697 gene products]
    - GO:0003824 : catalytic activity [51856 gene products]
    - GO:0030188 : chaperone regulator activity [73 gene products]
    - GO:0042056 : chemoattractant activity [14 gene products]
    - GO:0045499 : chemorepellent activity [9 gene products]
    - GO:0030234 : enzyme regulator activity [2370 gene products]
    - GO:0016530 : metallochaperone activity [47 gene products]
    - GO:0060089 : molecular transducer activity [7873 gene products]
    - GO:0003774 : motor activity [527 gene products]
    - GO:0045735 : nutrient reservoir activity [49 gene products]
    - GO:0031386 : protein tag [18 gene products]
    - GO:0005198 : structural molecule activity [4324 gene products]
    - GO:0030528 : transcription regulator activity [10429 gene products]
    - GO:0045182 : translation regulator activity [893 gene products]
    - GO:0005215 : transporter activity [10583 gene products]
Gene Ontology  Top level (CC)

- all : all [250418 gene products]
  - GO:0008150 : biological_process [166605 gene products]
  - GO:0005575 : cellular_component [169814 gene products]
    - GO:0005623 : cell [111086 gene products]
    - GO:0044464 : cell part [111049 gene products]
    - GO:0031975 : envelope [3316 gene products]
    - GO:0031012 : extracellular matrix [573 gene products]
    - GO:0044420 : extracellular matrix part [292 gene products]
    - GO:0005576 : extracellular region [5001 gene products]
    - GO:0044421 : extracellular region part [3365 gene products]
    - GO:0032991 : macromolecular complex [14668 gene products]
    - GO:0031974 : membrane-enclosed lumen [5290 gene products]
    - GO:0043226 : organelle [79653 gene products]
    - GO:0044422 : organelle part [16645 gene products]
    - GO:0055044 : symplast [3 gene products]
    - GO:0045202 : synapse [454 gene products]
    - GO:0044456 : synapse part [210 gene products]
    - GO:0019012 : virion [227 gene products]
    - GO:0044423 : virion part [186 gene products]
    - GO:0003674 : molecular_function [168558 gene products]
Gene Ontology  Top level (BP)

- all : all [250418 gene products]
  - GO:0008150 : biological_process [166605 gene products]
    - GO:0022610 : biological adhesion [1586 gene products]
    - GO:0065007 : biological regulation [31031 gene products]
    - GO:0001906 : cell killing [177 gene products]
    - GO:0009987 : cellular process [79087 gene products]
    - GO:0032502 : developmental process [19678 gene products]
    - GO:0051234 : establishment of localization [15270 gene products]
    - GO:0040007 : growth [4139 gene products]
    - GO:0002376 : immune system process [2517 gene products]
    - GO:0051179 : localization [17811 gene products]
    - GO:0040011 : locomotion [1251 gene products]
    - GO:0008152 : metabolic process [61127 gene products]
    - GO:0051704 : multi-organism process [4780 gene products]
    - GO:0032501 : multicellular organismal process [20567 gene products]
    - GO:0048519 : negative regulation of biological process [5037 gene products]
    - GO:0043473 : pigmentation [235 gene products]
    - GO:0048518 : positive regulation of biological process [6585 gene products]
    - GO:0050789 : regulation of biological process [28645 gene products]
    - GO:0000003 : reproduction [6343 gene products]
    - GO:0022414 : reproductive process [3535 gene products]
    - GO:0050896 : response to stimulus [16487 gene products]
    - GO:0048511 : rhythmic process [404 gene products]
    - GO:0016032 : viral reproduction [536 gene products]
Gene Ontology Examples

- all: all [250,418 gene products]
  - GO:0003674: molecular_function [168,558 gene products]
  - GO:0003824: catalytic activity [51,856 gene products]
  - GO:0016740: transferase activity [15,763 gene products]
    - GO:0016772: transferase activity, transferring phosphorus-containing groups
  - GO:0016301: kinase activity [60,931 gene products]
    - GO:0004672: protein kinase activity [3,504 gene products]
      - GO:0004712: protein serine/threonine/tyrosine kinase activity
    - GO:0004708: MAP kinase kinase activity
  - GO:0016773: phosphotransferase activity, alcohol group as acceptor
  - GO:0016773: phosphotransferase activity, alcohol group as acceptor
  - GO:0004672: protein kinase activity [3,504 gene products]
  - GO:0004712: protein serine/threonine/tyrosine kinase activity
    - GO:0004708: MAP kinase kinase activity
RxNorm
RxNorm Characteristics (1)

- Current version: June 2, 2007 (monthly releases)
- Type: Controlled terminology
- Domain: Drug names
- Developer: NLM
- Funding: NLM
- Availability
  - Publicly available: Yes*
  - Repositories: UMLS
- URL: http://www.nlm.nih.gov/research/umls/rxnorm/
RxNorm Characteristics (2)

◆ Number of
  ● Concepts: 93k
  ● Terms: 105k

◆ Major organizing principles:
  ● Generic vs. brand
  ● Combinations of Ingredient / Form / Dose
  ● No hierarchical structure
  ● Links to all major US drug information sources
  ● No clinical information

◆ Formalism: UMLS RRF format
RxNorm Normalized form

<table>
<thead>
<tr>
<th>Strength</th>
<th>Ingredient</th>
<th>Dose form</th>
</tr>
</thead>
<tbody>
<tr>
<td>4mg/ml</td>
<td>Fluoxetine</td>
<td>Oral Solution</td>
</tr>
</tbody>
</table>

Semantic clinical drug component

Semantic clinical drug form

Semantic clinical drug
Rx Norm  Generic vs. Brand

◆ Generic
  ● Ingredient (IN)
  ● Clinical drug form (SCDF)
  ● Clinical drug component (SCDC)
  ● Clinical drug (SCD)

◆ Brand
  ● Brand name (BN)
  ● Branded drug form (SBDF)
  ● Branded drug component (SBDC)
  ● Branded drug (SBD)

交易名称_of
RxNorm  Relations among drug entities
Medical Subject Headings (MeSH)
MeSH Characteristics (1)

- Current version: 2008 (yearly releases)
- Type: Thesaurus / Controlled vocabulary
- Domain: Biomedicine
- Developer: NLM
- Funding: NLM (Library Operations)

Availability
- Publicly available: Yes
- Repositories: UMLS / NCBO BioPortal

URL: http://www.nlm.nih.gov/mesh/
MeSH Characteristics (2)

- **Number of**
  - Terms: 7.5 per descriptor

- **Major organizing principles:**
  - Descriptor + entry terms
    (also: Qualifiers, Supplementary concepts)
  - Thesaurus relations (RB/RN/RO)

- **Formalism:** Thesaurus / Proprietary XML DTD
MeSH Top level

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Natural Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology, Industry, Agriculture [J]
11. Humanities [K]
12. Information Science [L]
13. Named Groups [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographicals [Z]
### MeSH Example (terms)

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Hydrocortisone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>D04.808.745.745.654.600</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D06.472.040.585.353.476</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D06.472.040.585.478.392</td>
</tr>
</tbody>
</table>

**Scope Note**

The main glucocorticoid secreted by the **ADRENAL CORTEX**. Its synthetic counterpart is used, either as an injection or topically, in the treatment of inflammation, allergy, collagen diseases, asthma, adrenocortical deficiency, shock, and some neoplastic conditions.

<table>
<thead>
<tr>
<th>Entry Term</th>
<th>11-Epicortisol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry Term</td>
<td>Cortifair</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortril</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Epicortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Hydrocortisone, (11 alpha)-Isomer</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Hydrocortisone, (9 beta, 10 alpha, 11 alpha)-Isomer</td>
</tr>
</tbody>
</table>
MeSH Example (hierarchies)

- Polycyclic Compounds
  - Steroids
    - Pregnanes
      - Pregnenes
        - Pregnenediones
  - Hormones, Hormone Substitutes, and Hormone Antagonists
    - Hormones
      - Adrenal Cortex Hormones
        - Hydroxycorticosteroids
          - 11-Hydroxycorticosteroids
  - Chemicals and Drugs
    - Hydrocortisone
NCI thesaurus  Characteristics (1)

- Current version: 08.04d (~monthly releases)
- Type: Controlled terminology / ontology
- Domain: Cancer
- Developer: NCI Center for Bioinformatics
- Funding: NCI
- Availability
  - Publicly available: Yes
  - Repositories: UMLS / OBO / NCBO BioPortal
- URL: http://nciterms.nci.nih.gov/
NCI thesaurus Characteristics (2)

- **Number of**
  - Concepts: 58,868 (2007_05E)
  - Terms: 2.68 per concept

- **Major organizing principles:**
  - Subsumption hierarchy
  - Rich set of associative relationships
  - Small proportion of defined concepts (many primitives)
  - Links to many external resources

- **Formalism:** OWL Lite
NCI_thesaurus  Top level

NCI_Thesaurus Taxonomy

- Abnormal Cell
- Activity
- Anatomic Structure, System, or Substance
- Biochemical Pathway
- Biological Process
- Chemotherapy Regimen or Agent Combination
- Conceptual Entity
- Diagnostic, Therapeutic, and Research Equipment
- Diagnostic or Prognostic Factor
- Disease, Disorder or Finding
- Drug, Food, Chemical or Biomedical Material
- Experimental Organism Anatomical Concept
- Experimental Organism Diagnosis
- Gene
- Gene Product
- Molecular Abnormality
- NCI Administrative Concept
- Organism
- Property or Attribute
- Retired Concept
## NCI thesaurus Example

### Concept Details

**Identifiers:**
- **name:** Prostate_Adenocarcinoma
- **code:** C2919

### Relationships to other concepts:
- **Disease Has Finding**
  - Invasive Lesion
- **Disease Has Abnormal Cell**
  - Adenocarcinoma Cell
- **Disease Has Normal Tissue Origin**
  - Prostatic Epithelium
- **Disease May Have Finding**
  - Serum Prostate Specific Antigen Increased
- **Disease Has Finding**
  - Carcinomatous Component Present
- **Disease Excludes Abnormal Cell**
  - Neoplastic Smooth Muscle Cell
- **Disease Excludes Abnormal Cell**
  - Malignant Squamous Cell
- **Disease Has Primary Anatomic Site**
  - Prostate Gland
- **Disease Has Associated Anatomic Site**
  - Male Reproductive System
- **Disease Excludes Abnormal Cell**
  - Malignant Stromal Cell
- **Disease Has Associated Anatomic Site**
  - Prostate Gland
- **Disease Has Normal Cell Origin**
  - Epithelial Cell

### Superconcepts:
- Adenocarcinoma
- Common Carcinoma
- Invasive Prostate Carcinoma

### Subconcepts:
- Acinar Prostate Adenocarcinoma
- Metastatic Prostatic Adenocarcinoma
- Moderately Differentiated Prostate Adenocarcinoma
- Poorly Differentiated Prostate Adenocarcinoma
- Prostate Adenocarcinoma with Focal Neuroendocrine Differentiation
- Prostate Ductal Adenocarcinoma
- Stage III Prostate Adenocarcinoma
- Stage II Prostate Adenocarcinoma
- Stage I Prostate Adenocarcinoma
- Well Differentiated Prostate Adenocarcinoma
Unified Medical Language System (UMLS)
UMLS Characteristics (1)

- Current version: 2008AA (2-3 annual releases)
- Type: Terminology integration system
- Domain: Biomedicine
- Developer: NLM
- Funding: NLM (intramural)
- Availability
  - Publicly available: Yes* (cost-free license required)
  - Repositories: UMLS
UMLS Characteristics (2)

- **Number of**
  - Concepts: 1.5M (2008AA)
  - Terms: ~6M

- **Major organizing principles (Metathesaurus):**
  - Concept orientation
  - Source transparency
  - Multi-lingual through translation

- **Formalism:** Proprietary format (RRF)
UMLS: Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Other subdomains
- SNOMED CT
- OMIM
- Biomedical literature
- MeSH
- NCBI Taxonomy
- Model organisms
- GO
- FMA
- Genome annotations
- Anatomy

Lister Hill National Center for Biomedical Communications
Addison’s Disease: Concept

An adrenal disease characterized by the progressive destruction of the adrenal cortex, resulting in insufficient production of aldosterone and hydrocortisone. Clinical symptoms include anorexia; nausea; weight loss; muscle weakness; and hyperpigmentation of the skin due to increase in circulating levels of ACTH precursor hormone which stimulates melanocytes.
**Metathesaurus Concepts** (2007AB)

- **Concept** (~1.4M) CUI
  - Set of synonymous concept names
- **Term** (~5.3M) LUI
  - Set of normalized names
- **String** (~5.9M) SUI
  - Distinct concept name
- **Atom** (~7.2M) AUI
  - Concept name in a given source

<table>
<thead>
<tr>
<th>CUI</th>
<th>Name</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0066000</td>
<td>Headache</td>
<td>(MeSH)</td>
</tr>
<tr>
<td>A0065992</td>
<td>Headache</td>
<td>(ICD-10)</td>
</tr>
<tr>
<td>A0066007</td>
<td>Headaches</td>
<td>(MedDRA)</td>
</tr>
<tr>
<td>A12003304</td>
<td>Headaches</td>
<td>(OMIM)</td>
</tr>
<tr>
<td>L0018681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0540936</td>
<td>Cephalodynia</td>
<td>(MeSH)</td>
</tr>
<tr>
<td>S0475647</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0380797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0018681</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lister Hill National Center for Biomedical Communications
<table>
<thead>
<tr>
<th>Name</th>
<th>Scope</th>
<th># concepts</th>
<th>Median</th>
<th>Subs. Hier</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED CT</td>
<td>Clinical medicine (patient records)</td>
<td>310,314</td>
<td>2</td>
<td>yes</td>
<td>July 31, 2007</td>
</tr>
<tr>
<td>LOINC</td>
<td>Clinical observations and laboratory tests</td>
<td>46,406</td>
<td>3</td>
<td>no</td>
<td>Version 2.21 (no “natural language” names)</td>
</tr>
<tr>
<td>FMA</td>
<td>Human anatomical structures</td>
<td>~72,000</td>
<td>?</td>
<td>yes</td>
<td>(not yet in the UMLS)</td>
</tr>
<tr>
<td>RxNorm</td>
<td>Standard names for prescription drugs</td>
<td>93,426</td>
<td>1</td>
<td>no</td>
<td>Aug. 31, 2007</td>
</tr>
<tr>
<td>NCI Thesaurus</td>
<td>Cancer research, clinical care, public information</td>
<td>58,868</td>
<td>2</td>
<td>yes</td>
<td>2007_05E</td>
</tr>
<tr>
<td>ICD-10</td>
<td>Diseases and conditions (health statistics)</td>
<td>12,318</td>
<td>1</td>
<td>no</td>
<td>1998 (tabular)</td>
</tr>
<tr>
<td>MeSH</td>
<td>Biomedicine (descriptors for indexing the literature)</td>
<td>24,767</td>
<td>5</td>
<td>no</td>
<td>Aug. 27, 2007</td>
</tr>
<tr>
<td>UMLS .</td>
<td>Terminology integration in the life sciences</td>
<td>1,4 M</td>
<td>2</td>
<td>n/a</td>
<td>2007AC (English only)</td>
</tr>
<tr>
<td>Monday, June 9</td>
<td>Introduction to Biomedical Ontologies</td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
<td>Biomedical Ontologies - Content and structure - Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Biomedical Ontologies in Action

A Functional Perspective on Biomedical Ontologies

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Functional perspective
  ● What are they for (vs. what are they)?
◆ “High-impact” biomedical ontologies
◆ 3 major categories of use
  ● Knowledge management (indexing and retrieval of data and information, access to information, mapping among ontologies)
  ● Data integration, exchange and semantic interoperability
  ● Decision support and reasoning (data selection and aggregation, decision support, natural language processing applications, knowledge discovery).
Knowledge management
Knowledge management

Annotating data and resources
Terminology in ontology

- Ontology as a source of vocabulary
  - List of names for the entities in the ontology (ontology vs. terminology)
- Most ontologies have some sort of terminological component
  - Exceptions: GALEN, LOINC
- Not all surface forms represented
  - Often insufficient for NLP applications
  - Large variation in number of terms per concept across ontologies
Annotating data

- Gene Ontology
  - Functional annotation of gene products in several dozen model organisms
- Various communities use the same controlled vocabularies
- Enabling comparisons across model organisms
- Annotations
  - Assigned manually by curators
  - Inferred automatically (e.g., from sequence similarity)
# GO Annotations for Aldh2 (mouse)

<table>
<thead>
<tr>
<th>Category</th>
<th>Classification Term</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Function</td>
<td><em>aldehyde dehydrogenase (NAD) activity</em></td>
<td>IEA</td>
</tr>
<tr>
<td>Molecular Function</td>
<td><em>oxidoreductase activity</em></td>
<td>IEA</td>
</tr>
<tr>
<td>Molecular Function</td>
<td><em>oxidoreductase activity</em></td>
<td>IEA</td>
</tr>
<tr>
<td>Cellular Component</td>
<td><em>mitochondrion</em></td>
<td>IDA</td>
</tr>
<tr>
<td>Biological Process</td>
<td><em>metabolic process</em></td>
<td>IEA</td>
</tr>
<tr>
<td>Biological Process</td>
<td><em>oxidation reduction</em></td>
<td>IEA</td>
</tr>
</tbody>
</table>

GO ALD4 in Yeast

GO Annotations

Molecular Function
Manually curated

Biological Process
Manually curated

Cellular Component
Manually curated
High-throughput

All ALD4 GO evidence and references

View Computational GO annotations for ALD4

- aldehyde dehydrogenase (NAD) activity (IDA, IMP, ISS)
- aldehyde dehydrogenase [NAD(P)+] activity (IDA)

- ethanol metabolic process (IMP)

- mitochondrial nucleoid (IDA)
- mitochondrion (IMP, ISS)
- mitochondrion (IDA)

http://db.yeastgenome.org/
### GO Annotations for ALDH2 (Human)

<table>
<thead>
<tr>
<th>Function</th>
<th>Database</th>
<th>Identifier</th>
<th>Code</th>
<th>Type</th>
<th>Pub ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0016491 oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR015590</td>
<td>UniProt 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0016491 oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR016160</td>
<td>UniProt 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0016491 oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR016162</td>
<td>UniProt 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0016491 oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR016161</td>
<td>UniProt 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0016491 oxidoreductase activity</td>
<td>spkw</td>
<td>EA</td>
<td>KW-0560</td>
<td>UniProt 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0034029 aldehyde dehydrogenase (NAD) activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0034030 aldehyde dehydrogenase [NAD(P)+] activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0039055 electron carrier activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0034029 aldehyde dehydrogenase (NAD) activity</td>
<td>enzyme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[http://www.ebi.ac.uk/GOA/](http://www.ebi.ac.uk/GOA/)
Indexing the biomedical literature

◆ MeSH
  ● Used for indexing and retrieval of the biomedical literature (MEDLINE)

◆ Indexing
  ● Performed manually by human indexers
    ▪ With help of semi-automatic systems (suggestions)
      e.g., Indexing Initiative at NLM
  ● Automatic indexing systems
Free cortisol in sepsis and septic shock.


Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@kku.fi

BACKGROUND: Severe sepsis activates the hypothalamo-pituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according to the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. RESULTS: Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ($r = 0.90$, $P < 0.001$). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock (728 +/- 386 nmol/L vs 793 +/- 439 nmol/L, $P = 0.44$). Nonsurvivors had higher calculated serum free (209 +/- 151 nmol/L) and total (980 +/- 458 nmol/L) cortisol concentrations than survivors (119 +/- 111 nmol/L, $P = 0.002$, and 704 +/- 383 nmol/L, $P = 0.002$). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%.

CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.
MeSH MEDLINE indexing

MeSH Terms:
- Adrenal Cortex Function Tests
- Adrenal Insufficiency/blood*
- Adrenal Insufficiency/drug therapy
- Adrenal Insufficiency/mortality
- Adult
- Biological Markers/blood
- Female
- Finland/epidemiology
- Hospital Mortality
- Humans
- Hydrocortisone/blood*
- Hydrocortisone/therapeutic use
- Kaplan-Meiers Estimate

- Male
- Predictive Value of Tests
- Prospective Studies
- Sepsis/blood*
- Sepsis/drug therapy
- Sepsis/mortality
- Severity of Illness Index
- Shock, Septic/blood*
- Shock, Septic/drug therapy
- Shock, Septic/mortality
- Treatment Outcome

Substances:
- Biological Markers
- Hydrocortisone
MeSH  MEDLINE indexing

Replacement therapy for Addison's disease: recent developments.

Lovás K, Husebye ES.

University of Bergen, Institute of Medicine, Section of Endocrinology, 5021 Bergen, Norway. Kristian.lovas@helse-bergen.no

BACKGROUND: The hormone deficiencies in Addison's disease (primary adrenal insufficiency) are conventionally treated with oral glucocorticoid and mineralocorticoid replacement but the available therapies do not restore the physiological hormone levels and biorhythm. Despite such treatment these patients self-report impaired health-related quality of life (HRQoL) and recent research has indicated increased mortality. OBJECTIVE/METHODS: We review the literature and recent developments in replacement therapy. RESULTS/CONCLUSION: Patients with Addison's disease require mineralocorticoid replacement, i.e., fludrocortisone 0.05 - 0.20 mg once daily. Starting doses of glucocorticoids should be 15 - 20 mg for hydrocortisone or 20 - 30 mg for cortisol acetate, divided into two or three doses, and preferentially weight-adjusted. There are indications that the synthetic glucocorticoids have undesirable metabolic long-term effects, which make them less suitable as first-line treatment. Timed-release hydrocortisone tablets and continuous subcutaneous hydrocortisone infusion are promising new treatment modalities. Studies of replacement with the adrenal androgen dehydroepiandrosterone (DHEA) in adrenal failure have shown inconsistent benefit on HRQoL. DHEA, or possibly testosterone replacement is likely to be beneficial for selected groups of patients with Addison's disease but this remains to be shown. We here give our opinion of the best treatment and future direction of research in this area.
MeSH MEDLINE indexing

MeSH Terms:
- Addison Disease/blood
- Addison Disease/drug therapy*
- Androgens/administration & dosage*
- Androgens/therapeutic use
- Dosage Forms
- Drug Administration Routes
- Drug Administration Schedule
- Glucocorticoids/administration & dosage*
- Glucocorticoids/adverse effects
- Glucocorticoids/blood
- Glucocorticoids/deficiency
- Hormone Replacement Therapy*
- Humans
- Mineralocorticoids/administration & dosage*
- Mineralocorticoids/adverse effects
- Mineralocorticoids/blood
- Mineralocorticoids/deficiency
- Quality of Life
- Treatment Outcome

Substances:
- Androgens
- Dosage Forms
- Glucocorticoids
- Mineralocorticoids
ICD9-CM Coding clinical data

◆ ICD9-CM
  - Used for coding clinical data e.g., for billing purposes

◆ Other uses of ICD
  - Morbidity and mortality reporting worldwide
Knowledge management

Accessing biomedical information
Resources for biomedical search engines

- Synonyms
- Hierarchical relations
- High-level categorization
- Co-occurrence information
- Translation
MeSH “synonyms” MEDLINE retrieval

- MeSH entry terms
  - Used as equivalent terms for retrieval purposes
  - Not always synonymous
- Increase recall without hurting precision

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Addison Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry Term</td>
<td>Addison's Disease</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Primary Adrenal Insufficiency</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Primary Adrenocortical Insufficiency</td>
</tr>
</tbody>
</table>
MeSH “synonyms” MEDLINE retrieval

Query Translation:

"addison disease"[MeSH Terms] OR ("addison"[All Fields] AND "disease"[All Fields]) OR "addison disease"[All Fields] OR ("primary"[All Fields] AND "hypoadrenalism"[All Fields]) OR "primary hypoadrenalism"[All Fields]
MeSH hierarchies  MEDLINE retrieval

◆ MeSH “explosion”
  - Search for a given MeSH term and all its descendants
  - A search on Adrenal insufficiency also retrieves articles indexed with Addison disease

Adrenal insufficiency

Addison disease
Items 1 - 20 of 8994


Co-indexing

http://www.gopubmed.com/

? cox-2

Lister Hill National Center for E
Knowledge management

Mapping across biomedical ontologies
Reusing information

- Clinical information coded with SNOMED CT
  - Mapped to ICD9-CM and CPT for billing purposes
  - Mapped to ICD-O for epidemiology purposes
- Existing mapping tables created by terminology developers as an incentive to use SNOMED CT
Reusing tools

- For noun phrases extracted from medical texts, map to UMLS concepts [Aronson & al., AMIA, 2000]
- Then, select from the MeSH vocabulary the concepts that are the most closely related to the original concepts
Terminology integration systems

- Terminology integration systems (UMLS, RxNorm) help bridge across vocabularies

- Uses
  - Information integration
  - Ontology alignment
  - Medication reconciliation
Integrating subdomains

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

OMIM

... 

UMLS

MeSH

Biomedical literature

Model organisms

NCBI Taxonomy

GO

FMA

Anatomy

Genome annotations

Lister Hill National Center for Biomedical Communications
Integrating subdomains

Clinical repositories

Genetic knowledge bases

Biomedical literature

Model organisms

Anatomy

Genome annotations

Other subdomains
Trans-namespacer integration

Addison's disease (363732003)

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

UMLS C0001403

Biomedical literature

Addison Disease (D000224)

NCBI Taxonomy

Model organisms

FMA

GO

Anatomy

Genome annotations

Lister Hill National Center for Biomedical Communications
Data integration, exchange and semantic interoperability
Data integration, exchange and semantic interoperability

Information exchange and semantic operability
“Standards”

◆ Ontologies help standardize patients data
  ● Facilitate the exchange of data across institutions
  ● Help connect “islands of data” (silos)

◆ LOINC
  ● Exchange of laboratory data
  ● In conjunction with HL7 messaging
Semantic interoperability projects  BRIDG

◆ Biomedical Research Integrated Domain Group
  ● Information model for clinical research
  ● Interoperability between clinical trials information systems
  ● Ontologies provide value sets to the information model
Semantic interoperability projects  CDA

◆ Clinical Document Architecture (CDA R2)
  • Formal representation of clinical statements
    ■ Clinical observations
    ■ Medication administration
    ■ Adverse events
  • Associate an information model (HL7 RIM) with terminologies (LOINC, SNOMED CT, RxNorm)
Semantic interoperability projects  caCORE

- Cancer Common Ontologic Representation Environment
  - Infrastructure developed to support an interoperable biomedical information system for cancer research
  - Uses the NCI Thesaurus as a component
Data integration, exchange and semantic interoperability

*Information and data integration*
Approaches to data integration

◆ Warehousing
  ● Sources to be integrated are transformed into a common format and converted to a common vocabulary
  ● Normalization through ontologies (e.g., GO annotations)

◆ Mediation
  ● Local schema (of the sources)
  ● Global schema (in reference to which the queries are made)
  ● Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)
Ontologies and integration

- Terminology integration systems help bridge across terminologies and the domains they represent
- Mappings across ontologies enable the integration of namespaces in the Semantic Web
Trans-namespaces integration

Addison's disease (363732003)

Clinical repositories

Genetic knowledge bases

Other subdomains

UMLS C0001403

Biomedical literature

Addison Disease (D000224)

Anatomy annotations

Model organisms

Genome annotations

Lister Hill National Center for Biomedical Communications 196 196
Decision support and reasoning
Data selection

- The structure of biomedical ontologies helps define groups of values from a high-level value
  - Vs. enumerating all possible values
- Useful for data selection in clinical studies
- ICD is used pervasively for this purpose
  - E.g., Study on supraventricular tachycardia (SVT), based on 2 high-level ICD codes
- Similarity with the definition of value sets for use in the information model
Data aggregation

- Ontologies help partition/aggregate data in data analysis
  - Clinical studies: Study a variable in groups of patients corresponding to the top level categories in ICD
  - Biology studies: Functional characterization of gene expression signatures with high-level concepts from the Gene Ontology
    - Recent trend: co-clustering
Decision support

◆ Clinical decision support
  • Ontologies help normalize the vocabulary and increase the recall of rules
  • Ontologies provide some domain knowledge and make it possible to create high-level rules (e.g., for a class of drugs rather than for each drug in the class)

◆ Other forms of decision support
  • Based on automatic reasoning services for OWL ontologies (e.g., grading gliomas with NCIt)
Natural language processing applications

- Ontologies provide background domain knowledge for NLP applications
  - Question answering
  - Document summarization
  - Literature-based discovery

- The UMLS is often used, but other specific resources have been developed
Knowledge discovery

- By standardizing the vocabulary in a given domain, ontologies are enabling resources for knowledge discovery through data mining
- Less frequently, the structure of the ontology is leveraged by data mining algorithms
- Example of available datasets
  - ICD-coded clinical data (in conjunction with non-clinical information, e.g., environmental data)
  - Annotation of gene products to the GO (function prediction)
Barriers to usability of biomedical ontologies
Availability

◆ Many ontologies are freely available
◆ The UMLS is freely available for research purposes
  ● Cost-free license required
◆ Licensing issues can be tricky
  ● SNOMED CT is freely available in member countries of the IHTSDO
◆ Being freely available
  ● Is a requirement for the Open Biomedical Ontologies (OBO)
  ● Is a de facto prerequisite for Semantic Web applications
Discoverability

- Ontology repositories
  - UMLS: 143 source vocabularies
    (biased towards healthcare applications)
  - NCBO BioPortal: ~100 ontologies
    (biased towards biological applications)
  - Limited overlap between the two repositories
- Need for discovery services
Formalism

◆ Several major formalism
   - Web Ontology Language (OWL) – NCI Thesaurus
   - OBO format – most OBO ontologies
   - UMLS Rich Release Format (RRF) – UMLS, RxNorm

◆ Conversion mechanisms
   - OBO to OWL
   - LexGrid (import/export to LexGrid internal format)
Ontology integration

◆ *Post hoc* integration, form the bottom up
  - UMLS approach
  - Integrates ontologies “as is”, including legacy ontologies
  - Facilitates the integration of the corresponding datasets

◆ Coordinated development of ontologies
  - OBO Foundry approach
  - Ensures consistency *ab initio*
  - Excludes legacy ontologies
Quality

◆ Quality assurance in ontologies is still imperfectly defined
  ● Difficult to define outside a use case or application

◆ Several approaches to evaluating quality
  ● Collaboratively, by users (Web 2.0 approach)
    ■ Marginal notes enabled by BioPortal
  ● Centrally, by experts
    ■ OBO Foundry approach

◆ Important factors besides quality
  ● Governance
  ● Installed base / Community of practice
# Agenda

<table>
<thead>
<tr>
<th>Date</th>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monday, June 9</strong></td>
<td>Introduction to Biomedical Ontologies</td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
<td>Biomedical Ontologies - Content and structure - Function</td>
</tr>
<tr>
<td><strong>Tuesday, June 10</strong></td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td><strong>Wednesday, June 11</strong></td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Interfaces to Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Graphical interfaces
  ● UMLS Knowledge Source Server
  ● NCBO BioPortal
  ● NCI Thesaurus
  ● MeSH browser
  ● Foundational Model of Anatomy (FMA) Explorer
  ● Gene Ontology AmiGO
  ● ICD-10 online
  ● RxNav (RxNorm)
  ● […]

◆ Application Programming Interfaces
Graphical interfaces
UMLS Knowledge Source Server

http://umlsks.nlm.nih.gov/

Lister Hill National Center for Biomedical Communications
UMLS Semantic Navigator


Lister Hill National Center for Biomedical Communications
**NCBO BioPortal**

### Ontologies

<table>
<thead>
<tr>
<th>Name</th>
<th>Format</th>
<th>Current Version</th>
<th>Content Location</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Traditional Medicine</td>
<td>OBO</td>
<td>1.0.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Amino Acid</td>
<td>OWL Full</td>
<td>1.2</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Amphibian gross anatomy</td>
<td>OBO</td>
<td>1.7</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Animal natural history and life history</td>
<td>Protege</td>
<td>See Remote Site</td>
<td>Remote</td>
<td></td>
</tr>
<tr>
<td>Basic Vertebrate Anatomy</td>
<td>OWL Full</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biological imaging methods</td>
<td>OBO</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biological process</td>
<td>OBO</td>
<td>1.208</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biomedical Resource Ontology</td>
<td>OWL Lite</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>BIRNLex</td>
<td>OWL DL</td>
<td>1.3.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
</tbody>
</table>

http://www.bioontology.org/tools/portal/bioportal.html

Lister Hill National Center for Biomedical Communications
<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Prostatic Neoplasms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>C04 588 945 440 770</td>
</tr>
<tr>
<td>Tree Number</td>
<td>C12 294 260 750</td>
</tr>
<tr>
<td>Tree Number</td>
<td>C12 294 565 625</td>
</tr>
<tr>
<td>Tree Number</td>
<td>C12 758 409 750</td>
</tr>
<tr>
<td>Annotation</td>
<td>coordinate IM with histological type of neoplasm (IM); note PROSTATIC ADENOMA see PROSTATIC HYPERPLASIA is also available</td>
</tr>
<tr>
<td>Scope Note</td>
<td>Tumors or cancer of the PROSTATE</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cancer of Prostate</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cancer of the Prostate</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Neoplasms, Prostate</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Neoplasms, Prostatic</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Prostate Cancer</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Prostate Neoplasms</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Prostatic Cancer</td>
</tr>
<tr>
<td>See Also</td>
<td>Prostate-Specific Antigen</td>
</tr>
<tr>
<td>See Also</td>
<td>Prostatic Hyperplasia</td>
</tr>
</tbody>
</table>
| Allowable Qualifiers | BL BS CF CH CI CL CN CO D
|                    | E D F G H I IM ME MI MO NU P
|                    | Q R S T U V W X Y Z            |
| Entry Version      | PROSTATIC NEOPL                |
| Date of Entry      | 19990101                       |
| Unique ID          | D011471                        |


Lister Hill National Center for Biomedical Communications
Foundational Model of Anatomy

http://sig.biostr.washington.edu/projects/fm/FME/

Lister Hill National Center for Biomedical Communications
Gene Ontology AmiGO

Term Information
Accession: GO:0008375
Ontology: molecular function
Definition: exact: GlcNAc transferase activity

Term Lineage
Switch to viewing term parents, siblings and children

Filter tree view
Filter Gene Product Counts
Data source: All
Species: All

View Options
Tree view
Set filters
Remove all filters

Actions...
Last action: Reset the tree
Graphical View
View in tree browser
Download...
RDO
RDF-XML
GraphViz dot

Back to top

http://www.geneontology.org/

Lister Hill National Center for Biomedical Communications
**Viral meningitis (disorder)**

| CONCEPT | | | | | |
| --- | --- | --- | --- | --- |
| Concept ID | Fully Specified Name | Concept Status | CTV3ID | SNOMED ID | Is Primitive |
| 58170007 | Viral meningitis (disorder) | Current (0) | xa9b5 | DE-30020 | Fully defined (0) |

| DESCRIPTIONS and SYNONYMS | | | | | |
| --- | --- | --- | --- | --- |
| Description ID | Term | Description Status | Description Type | Language Code | Initial Capital Status |
| 1 | 96672018 | Viral meningitis | Current (0) | Preferred (1) | en | Capitalization meaningless (0) |
| 2 | 96673011 | Aseptic meningitis, viral | Current (0) | Synonym (2) | en | Capitalization meaningless (0) |
| 3 | 96674017 | Abacterial meningitis | Current (0) | Synonym (2) | en | Capitalization meaningless (0) |
| 4 | 96677012 | Aseptic meningitis | Non-Current (1) | Synonym (2) | en | Capitalization meaningless (0) |
| 5 | 96676015 | Viral meningitis NOS | Non-Current (1) | Synonym (2) | en | Capitalization meaningless (0) |
| 6 | 96675016 | Viral meningitis, NEC | Non-Current (1) | Synonym (2) | en | Capitalization meaningless (0) |

| PARENTS | | | |
| --- | --- | --- |
| Concept ID | FSN for Parent Concept (This Concept IS A) | |
| 1 | 302810003 | Viral infections of the central nervous system (disorder) |
| 2 | 312216007 | Infective meningitis (disorder) |

http://www.it.usyd.edu.au/~hitru/sct/A1.cgi
SNOMED CT  CliniClue

www.cliniclue.com/

Lister Hill National Center for Biomedical Communications
ICD-10

Chapter IV

Endocrine, nutritional and metabolic diseases
(E00-E90)

Disorders of other endocrine glands
(E20-E35)

Excludes: galactorrhoea (N54.3)
gynaecomastia (N62)

E20 Hypoparathyroidism

Excludes: Di George's syndrome (D58.1)
postprocedural hypoparathyroidism (E09.2)
tetany NOS (R29.0)
transitory neonatal hypoparathyroidism (P71.4)

E20.0 Idiopathic hypoparathyroidism
E20.1 Pseudohypoparathyroidism
E20.8 Other hypoparathyroidism
E20.9 Hypoparathyroidism, unspecified
Parathyroid tetany

http://www.who.int/classifications/apps/icd/icd10online/

Lister Hill National Center for Biomedical Communications
RxNav (RxNorm)

Application Programming Interfaces
Application Programming Interface

- Expose resources in such a way that they can be integrated in programs
  - Programming “against” a resource
- Standard protocols for communication
  - Web services (SOAP, REST)
- Standard libraries for programming
- Focus on content, not message
UMLSKS Web Service API

  - Developer's Guide > Webservice Operations
- WSDL available
- API give access to all 3 knowledge sources
- Licensing issues
  - Granting ticket and Single-use tickets
ConceptIdGroup findCUIByNormString(ConceptIdNormStringRequest request);

**Argument:** ConceptIdNormStringRequest

This class contains the arguments that further restrict the behavior of the call.

- setCasTicket (String s)
  - Single-use ticket returned by the AuthorizationPort webservice
- setRelease (String s)
  - UMLS release of interest
- setSearchString (String s)
  - input search string
- setSABs (String[] array)
  - set of source abbreviations to search
- setLanguage (String s)
  - language restriction
- setIncludeSuppressibles (boolean b)
  - true if suppressible strings are included in the search
- setCVF (long l)
  - Bit flag for the content view to search

**Return:** ConceptIdGroup
Other APIs to terminology systems

- **NCBO BioPortal**
  
  [http://www.bioontology.org/docs/bioportal/development/web_services.html](http://www.bioontology.org/docs/bioportal/development/web_services.html)

- **OLS - Ontology Lookup Service**
  
  [http://www.ebi.ac.uk/ontology-lookup/WSDLDocumentation.do](http://www.ebi.ac.uk/ontology-lookup/WSDLDocumentation.do)

- **RxNorm**
  
Applications based on WS APIs

- **UMLSKS API**
  - UMLSKS

- **RxNorm API**
  - RxNav
  - MyMedicationList
<table>
<thead>
<tr>
<th></th>
<th>Monday, June 9</th>
<th>Tuesday, June 10</th>
<th>Wednesday, June 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction to Biomedical Ontologies</strong></td>
<td><strong>Design Principles, Formalisms and Tools for Biomedical Ontologies</strong></td>
<td><strong>Biomedical Ontologies</strong> - Content and structure - Function</td>
<td></td>
</tr>
<tr>
<td><strong>Interfaces to Biomedical Ontologies</strong></td>
<td><strong>Searching and Analyzing Biomedical Concepts</strong></td>
<td><strong>Contrasting Biomedical Ontologies</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Critical Analysis of Biomedical Ontologies</strong></td>
<td><strong>Extending Biomedical Ontologies</strong></td>
<td><strong>Using Biomedical Ontologies for Data Integration</strong></td>
<td></td>
</tr>
</tbody>
</table>
Searching and Analyzing Biomedical Concepts

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Exercise 1

◆ What are the Clinical Drug Components for Zyrtec? (RxNav)
Exercise 2

◆ What are the parts of the Aorta? (FMA)
Exercise 3

◆ What are the parents of Hodgkin’s disease in SNOMED CT?
  ● Try SNOMEDCTID: 118599009
◆ What is its associated morphology?
Exercise 4

- What are the various meanings of IL-2? (UMLS)
Exercise 5

◆ What are the pharmacologic actions of Zyrtec? (MeSH)
Exercise 6

What are some synonyms for Schwannoma? (NCI Thesaurus)
Solutions
Exercise 1

◆ What are the Clinical Drug Components for Zyrtec? (RxNav)
Exercise 2

- What are the parts of the Aorta? (FMA)

[Diagram of the Aorta with parts listed: Neural network of cardiovascular system, Systemic arterial tree, Aorta, Right coronary artery, Left coronary artery, Brachiocephalic artery, Common carotid artery, Ascending aorta, Arch of aorta, Descending aorta, Wall of aorta, Lumen of aorta]
Exercise 3

◆ What are the parents of Hodgkin’s disease in SNOMED DT?
  ● Try SNOMEDCTID: 118599009
◆ What is its associated morphology?

<table>
<thead>
<tr>
<th>CONCEPT</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>Fully Specified Name</td>
<td>Concept Status</td>
<td>CTV3ID</td>
<td>SNOMED ID</td>
<td>Is Primitive</td>
</tr>
<tr>
<td>118599009</td>
<td>Hodgkin’s disease (disorder)</td>
<td>Current (0)</td>
<td>B61..</td>
<td>DC-F1000</td>
<td>Fully defined (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARENTS</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>FSN for Parent Concept (This Concept IS A)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>118600007</td>
<td>Malignant lymphoma (disorder)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>FSN for Target Concept</td>
<td>Relationship Type</td>
<td>Values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>288526004</td>
<td>Episodicities (qualifier value)</td>
<td>Episodicity (attribute)</td>
<td>{New episode; Ongoing episode; Old episode; Undefined episodicity; Other episode RCGP; First episode}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Exercise 4

◆ What are the various meanings of IL-2? (UMLS)

![Metathesaurus Search](image)

**IL2 gene [C0879590]**
**Interleukin-2 [C0021756]**
**Recombinant Interleukin-2 [C1522405]**
**interleukin-2 binding [C1149229]**
Exercise 5

◆ What are the pharmacologic actions of Zyrtec? (MeSH)

| Pharm. Action | Anti-Allergic Agents |
| Pharm. Action | Histamine H1 Antagonists, Non-Sedating |
Exercise 6

◆ What are some synonyms for Schwannoma? (NCI Thesaurus)

Preferred_Name  Schwannoma
Semantic_Type   Neoplastic Process
Synonym         Neurilemmoma
Synonym         Neurinoma
Synonym         Schwannoma
Synonym         schwannoma
What to look for

- Search modalities
  - Spelling correction / auto-completion / normalization
  - Word combinations (AND/OR)
- Visualization
  - Graph vs. (forest of) trees
- Navigation
- What properties are displayed
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, June 9</td>
<td>Introduction to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Biomedical Ontologies - Content and structure - Function</td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Searching and Analyzing Biomedical Concepts</td>
</tr>
<tr>
<td></td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Extending Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Exercise #1

- Hodgkin’s disease
  - NCI Thesaurus
  - SNOMED CT
Exercise #2

- Prostate
  - FMA
  - SNOMED CT
Exercise #3

- Cetirizine
  - MeSH
  - SNOMED CT
Solutions
Solutions

Exercise #1
Exercise #1

◆ Hodgkin’s disease
  ● NCI Thesaurus
    ■ Using the NCI browser (EVS)
      http://nciterms.nci.nih.gov/
  ● SNOMED CT
    ■ Using the online browser from U. Sydney
Hodgkin’s disease in NCI (1)

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Version:</td>
<td>April 2008 (08.04d)</td>
</tr>
</tbody>
</table>

**Hodgkin Lymphoma**

<table>
<thead>
<tr>
<th>Identifiers:</th>
<th>Hodgkin_s_Lymphoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Hodgkin_s_Lymphoma</td>
</tr>
<tr>
<td>code</td>
<td>C9357</td>
</tr>
</tbody>
</table>

**Relationships to other concepts:**

<table>
<thead>
<tr>
<th>Disease_Has_Primary_Anatomic_Site</th>
<th>Hematopoietic and Lymphatic System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease_Has_Normal_Tissue_Origin</td>
<td>Lymphoid Tissue</td>
</tr>
<tr>
<td>Disease_Excludes_Normal_Cell_Origin</td>
<td>Myeloid Cell</td>
</tr>
<tr>
<td>Disease_Excludes_Normal_Cell_Origin</td>
<td>Plasma Cell</td>
</tr>
<tr>
<td>Disease_Has_Abnormal_Cell</td>
<td>Reed-Sternberg Cell</td>
</tr>
<tr>
<td>Disease_Has_Associated_Anatomic_Site</td>
<td>Hematopoietic and Lymphatic System</td>
</tr>
<tr>
<td>Disease_Has_Normal_Cell_Origin</td>
<td>Mature Lymphocyte</td>
</tr>
<tr>
<td>Disease_Has_Primary_Anatomic_Site</td>
<td>Lymphatic System</td>
</tr>
</tbody>
</table>

**Superconcepts**

- Common Hematopoietic Neoplasm
- Lymphoma
## Hodgkin’s disease in NCIt (1)

### Information about this concept:

<table>
<thead>
<tr>
<th>ALT_DEFINITION</th>
<th>A malignant disease of the lymphatic system that is characterized by painless enlargement of lymph nodes, the spleen, or other lymphatic tissue. It is sometimes accompanied by symptoms such as fever, weight loss, fatigue, and night sweats.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFINITION</td>
<td>A lymphoma, previously known as Hodgkin’s disease, characterized by the presence of Reed-Sternberg cells. There are two distinct subtypes: nodular lymphocyte predominant Hodgkin lymphoma and classical Hodgkin lymphoma. Hodgkin lymphoma has a bimodal age distribution, and involves primarily lymph nodes. Current therapy for Hodgkin lymphoma has resulted in an excellent outcome and cure for the majority of patients.</td>
</tr>
</tbody>
</table>

| ICD-0-3_Code   | 9650/3 |
| Preferred_Name | Hodgkin Lymphoma |
| Semantic_Type  | Neoplastic Process |
| Synonym        | HL |
| Synonym        | Hodgkin Lymphoma |
| Synonym        | Hodgkin's Disease |
| Synonym        | Hodgkin's Lymphoma |
| Synonym        | Hodgkin's disease |

| Unified Medical Language System Concept Identifier | C0019829 |
Comments on Hodgkin’s disease in NCIt (1)

◆ Search term: “Hodgkin’s disease”
  ● Not found, although “Hodgkin’s disease” is listed as a synonym
  ● Search on “hodgkin”, select “Hodgkin lymphoma”
◆ Parent classes
  ● Common hematopoietic neoplasm
    ■ Not an ontological category
    ■ Would be better represented through an associative relation (e.g., along the lines of “has_prevalence high prevalence”)
    ■ Isa overloading
Comments on Hodgkin’s disease in NCIt (2)

◆ Associative relations
  
  • For cancers, anatomy and morphology are foundational relations
  
  • Here
    
    ▶ Anatomy: Disease_Has_Primary_Anatomic_Site
    Hematopoietic and Lymphatic System

    ▶ Morphology: not directly represented
      (indirectly through Disease_Has_Normal_Cell_Origin
      Mature Lymphocyte)
Hodgkin’s disease in SNOMED CT (1)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Concept Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1185990009</td>
<td>Hodgkin's disease (disorder)</td>
<td>Current (0)</td>
</tr>
</tbody>
</table>

**Parents**

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Malignant lymphoma (disorder)</td>
</tr>
</tbody>
</table>

**Attributes**

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Target Concept</th>
<th>Relationship Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Episodicities (qualifier value)</td>
<td>Episodicity (attribute)</td>
</tr>
<tr>
<td>2</td>
<td>Hodgkin lymphoma - category</td>
<td>Associated morphology (attribute)</td>
</tr>
<tr>
<td>2</td>
<td>Hodgkin lymphoma - category</td>
<td>Associated morphology (attribute)</td>
</tr>
</tbody>
</table>
Hodgkin’s disease in SNOMED CT (2)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Concept Status</th>
<th>CTV3ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1185990009</td>
<td>Hodgkin's disease (disorder)</td>
<td>Current (0)</td>
<td>B61..</td>
</tr>
</tbody>
</table>

**Descriptions and Synonyms**

<table>
<thead>
<tr>
<th>Description ID</th>
<th>Term</th>
<th>Description Status</th>
<th>Description Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hodgkin's disease (clinical)</td>
<td>Current (0)</td>
<td>Preferred (1)</td>
</tr>
<tr>
<td>2</td>
<td>Malignant Hodgkin's lymphoma</td>
<td>Current (0)</td>
<td>Synonym (2)</td>
</tr>
<tr>
<td>3</td>
<td>HD - Hodgkin's disease</td>
<td>Current (0)</td>
<td>Synonym (2)</td>
</tr>
</tbody>
</table>
Comments on Hodgkin’s disease in SNOMED CT (1)

◆ Search term: “Hodgkin’s disease”
  ● Not found, although “Hodgkin’s disease” is listed as a synonym
    ■ Search result: “Hodgkin lymphoma, nodular sclerosis, grade 1 (morphologic abnormality)”
  ● Search on “lymphoma”, navigate down from “Malignant lymphoma”
  ● “hodgkin’s disease” is ambiguous
    ■ Hodgkin lymphoma, no ICD-O subtype (morphologic abnormality)
    ■ Hodgkin's disease (disorder)
  ● “Malignant lymphoma, Hodgkin's”
    ■ NB: lymphoma is always malignant

◆ Parent classes
  ● Malignant lymphoma (clinical) [OK]
Comments on Hodgkin’s disease in SNOMED CT (2)

◆ Associative relations

- For cancers, anatomy and morphology are foundational relations
- Here
  - Anatomy: not directly represented (indirectly through descendant concepts, e.g., Hodgkin's disease of intrathoracic lymph nodes)
  - Morphology: Associated morphology Hodgkin lymphoma - category
Hodgkin’s disease  NCIt vs. SNOMED CT (1)

◆ Shared synonyms: NCIt 1/2, SNOMED CT 1/3
  ● Hodgkin’s disease

◆ Shared relations
  ● Isa
    ■ NCIt: Lymphoma
      – Definition: “malignant (clonal) proliferation of B-lymphocytes or T-lymphocytes which involves the lymph nodes, bone marrow and/or extranodal sites. This category includes Non-Hodgkin lymphomas and Hodgkin lymphomas.”
    ■ SNOMED CT: Malignant lymphoma
    ■ Same UMLS concept (CUI: C0024299)
Hodgkin’s disease  NCIt vs. SNOMED CT (2)

- **Shared relations: Associative relations**
  - Anatomy
    - In NCIt, but not in SNOMED CT
  - Morphology
    - In SNOMED CT, but not in NCIt
      - Only indirectly, though cell type
  - Cell type
    - Only in NCIt
Solutions

Exercise #2
Exercise #2

◆ Prostate
  ● FMA
    ■ Using the Foundational Model Explorer
      http://sig.biostr.washington.edu/projects/fm/FME/
  ● SNOMED CT
    ■ Using the online browser from U. Sydney
Prostate in FMA (1)

Foundational Model Explorer

Search

Organ system subdivision
Organ
Solid organ
Parenchymatous organ
Lobular organ
Lung
Liver
Pancreas
Prostate
Testis

PART OF:

Genital system
Content of male pelvis
Set of male pelvic viscera
Set of pelvic viscera

Intraprostatic part of left ejaculatory duct
Prostatic stroma
Neural network of prostate
Vasculature of prostate
Prostate in FMA (2)

Foundational Model Explorer

PREFERRED NAME: Prostate

NON-ENGLISH EQUIVALENT:

<table>
<thead>
<tr>
<th>name</th>
<th>language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostata (Glandula prostatica)</td>
<td>Latin</td>
</tr>
<tr>
<td>Prostata</td>
<td>Latin</td>
</tr>
<tr>
<td>Próstata</td>
<td>Spanish</td>
</tr>
</tbody>
</table>

FMAID: 9600

Organ system subdivision
- Organ
  - Solid organ
    - Parenchymatous organ
  - Lobular organ
    - Lung
    - Liver
    - Pancreas
    - Prostate
  - Testis
  - Salivary gland
Comments on Prostate in FMA

◆ No synonyms in English
  ● Latin and Spanish synonyms

◆ Hierarchies
  ● Isa: Lobular organ
  ● Part_of: Set of pelvic viscera

◆ Associative relations
  ● Lymphatic drainage
  ● No spatial relations
### Prostate in SNOMED CT (1)

#### 435 results found for prostate:

<table>
<thead>
<tr>
<th>#</th>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Preferred Terms and Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9713002</td>
<td>Prostatitis (disorder)</td>
<td>Inflammation of prostate- Prostatitis [PT]- Prostatitis, NOS</td>
</tr>
<tr>
<td>2</td>
<td>11441004</td>
<td>Prostatism (disorder)</td>
<td>Prostatism [PT]- Prostatism, NOS</td>
</tr>
<tr>
<td>3</td>
<td>41216001</td>
<td>Prostatic structure (body structure)</td>
<td>Prostatic structure [PT]- Prostate- Prostate, NOS</td>
</tr>
<tr>
<td>4</td>
<td>181422007</td>
<td>Entire prostate (body structure)</td>
<td>Entire prostate [PT]- Prostate</td>
</tr>
</tbody>
</table>

#### CONCEPT

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>181422007</td>
<td>Entire prostate (body structure)</td>
</tr>
</tbody>
</table>

#### DESCRIPTIONS and SYNONYMS

<table>
<thead>
<tr>
<th>Description ID</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>280451017</td>
<td>Entire prostate</td>
</tr>
<tr>
<td>280452012</td>
<td>Prostate</td>
</tr>
</tbody>
</table>

#### PARENTS

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept IS A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male internal genital organ (body structure)</td>
</tr>
<tr>
<td>2</td>
<td>Prostatic structure (body structure)</td>
</tr>
<tr>
<td>3</td>
<td>Entire male genital organ (body structure)</td>
</tr>
</tbody>
</table>

Lister Hill National Center
# Prostate in SNOMED CT (2)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Target Concept</th>
<th>Relationship Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Entire viscus (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>2</td>
<td>Entire urinary tract (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>3</td>
<td>Entire male internal genitalia (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>4</td>
<td>Entire lower male genitourinary tract (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>5</td>
<td>Entire minor pelvis (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>6</td>
<td>Entire lower genitourinary tract (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>7</td>
<td>Entire male genital system (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>8</td>
<td>Entire abdomen (body structure)</td>
<td>Part of (attribute)</td>
</tr>
</tbody>
</table>
Comments on Prostate in SNOMED CT

◆ “Ambiguous” term
  • Entire prostate
  • Prostatic structure

◆ Structure-Entire-Part representation of anatomical entities in SNOMED CT
  • Reification of \textit{part\_of}
  • Enables mereological inference through \textit{isa} hierarchy
  • Not intuitive
Structure-Entire-Part (SEP) triples

- **S** – The entity or any of its parts
- **E** – The entire anatomical entity
- **P** – Any parts of the anatomical entity

[Schulz & al., 1997]
[Schulz & al., 1998]
[Bodenreider et al, 2006]
Mereological inference through *isa*

Diagram showing the relationships:
- Kidney structure
- Entire kidney
- Kidney part
  - Structure of layer of kidney
    - Entire cortex of kidney
  - Structure of region of kidney
    - Entire pole of kidney

*isa* and *part_of* relationships are shown with arrows.
FMA mapping goes to *Entire*

- **FMA**

- **SNOMED CT**

  - Kidney
  - Kidney *structure*
  - Kidney *part*
  - *Entire kidney*

*isa* relation:
- Kidney to Kidney *structure*
- Kidney *part* to *Entire kidney*

*part_of* relation:
- Kidney *part* to *Entire kidney*

Synonyms: Kidney

Lister Hill National Center for Biomedical Communications
Prostate  FMA vs. SNOMED CT

◆ Shared synonyms: FMA 1/1, SNOMED CT 1/2
  ● Prostate

◆ Shared relations
  ● Isa: no
    ■ FMA
      – Lobular organ
    ■ SNOMED CT
      – Prostatic structure
      – Male internal genital organ
      – Entire male genital organ
Prostate  FMA vs. SNOMED CT

% Shared relations

- Part of: almost
  - FMA
    - Genital system
    - Content of male pelvis
      - Set of male pelvic viscera
      - Set of pelvic viscera
  - SNOMED CT
    - Entire minor pelvis
    - Entire male genital system
    - ...

Lister Hill National Center for Biomedical Communications
Solutions

Exercise #3
Exercise #3

◆ Cetirizine
  ● MeSH
    ▪ Using the MeSH browser
  ● SNOMED CT
    ▪ Using the online browser from U. Sydney
Cetirizine in MeSH (1)

<table>
<thead>
<tr>
<th>Entry Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2-(4-((4-Chlorophenyl)phenylmethyl)-1-piperazinyl)ethoxy)acetic Acid</td>
<td></td>
</tr>
<tr>
<td>Alerlisin</td>
<td></td>
</tr>
<tr>
<td>Aliud Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Alpharma Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>AWD.pharma Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Azupharma Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Basics Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Cetalerg</td>
<td></td>
</tr>
<tr>
<td>Voltric</td>
<td></td>
</tr>
<tr>
<td>Wolff Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Wörwag Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Zetir</td>
<td></td>
</tr>
<tr>
<td>Zirtek</td>
<td></td>
</tr>
<tr>
<td>Zyrtec</td>
<td></td>
</tr>
</tbody>
</table>
Cetirizine in MeSH (2)

- Heterocyclic Compounds [D03]
  - Heterocyclic Compounds, 1-Ring [D03.383]
  - Piperazines [D03.383.606]
  - Hydroxyzine [D03.383.606.515]

- Cetirizine [D03.383.606.515.200]

| Pharm. Action | Anti-Allergic Agents |
| Pharm. Action | Histamine H1 Antagonists, Non-Sedating |
Comments on Cetirizine in MeSH

◆ 45 entry terms
  ● Various generic and brand names
  ● Chemical formula
  ● Code (P-071)

◆ Hierarchy
  ● *Isa*: Piperazines [chemistry]

◆ Pharmacologic action
  ● Anti-Allergic Agents
  ● Histamine H1 Antagonists, Non-Sedating
### 15 results found for cetirizine:

<table>
<thead>
<tr>
<th>#</th>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Preferred Terms and Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1086550000</td>
<td>Cetirizine (product)</td>
<td>Cetirizine [PT]</td>
</tr>
<tr>
<td>2</td>
<td>372523007</td>
<td>Cetirizine (substance)</td>
<td>Cetirizine [PT]</td>
</tr>
</tbody>
</table>

#### Cetirizine (substance)

<table>
<thead>
<tr>
<th>CONCEPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>372523007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DESCRIPTIONS and SYNONYMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description ID</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>1211057019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>372624008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHILDREN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>108656004</td>
</tr>
</tbody>
</table>
### Cetirizine in SNOMED CT (2)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Concept Status</th>
<th>CTV3ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>108655000</td>
<td>Cetirizine (product)</td>
<td>Current (0)</td>
<td>☐01Dq</td>
</tr>
</tbody>
</table>

#### DESCRIPTIONS and SYNONYMS

<table>
<thead>
<tr>
<th>Description ID</th>
<th>Term</th>
<th>Description Status</th>
<th>Description Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>173189012</td>
<td>Cetirizine</td>
<td>Current (0)</td>
<td>Preferred (1)</td>
</tr>
</tbody>
</table>

#### PARENTS

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept IS A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>340956006</td>
<td>Non-sedating antihistamine (product)</td>
</tr>
</tbody>
</table>

#### ATTRIBUTES

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Target Concept</th>
<th>Relationship Type</th>
<th>Parent</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>372523007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### CHILDREN

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Child Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>320818006</td>
<td>Cetirizine dihydrochloride 10mg tablet (product)</td>
</tr>
<tr>
<td>320820009</td>
<td>Cetirizine dihydrochloride 1mg/1mL sff liquid (product)</td>
</tr>
<tr>
<td>371746005</td>
<td>Cetirizine dihydrochloride 5mg tablet (product)</td>
</tr>
<tr>
<td>375571002</td>
<td>Cetirizine hydrochloride 5mg tablet (product)</td>
</tr>
<tr>
<td>375572009</td>
<td>Cetirizine hydrochloride 10mg tablet (product)</td>
</tr>
<tr>
<td>375573004</td>
<td>Cetirizine hydrochloride 5mg/5 mL syrup (product)</td>
</tr>
<tr>
<td>400482001</td>
<td>Cetirizine hydrochloride+pseudoephedrine hydrochloride (product)</td>
</tr>
<tr>
<td>409491005</td>
<td>Cetirizine hydrochloride 5mg chewable tablet (product)</td>
</tr>
<tr>
<td>409492003</td>
<td>Cetirizine hydrochloride 10mg chewable tablet (product)</td>
</tr>
</tbody>
</table>
Comments on Cetirizine in SNOMED CT

- Ambiguous term
  - Cetirizine (product)
  - Cetirizine (substance)
- Hierarchy
  - Isa: Non-sedating antihistamine (substance) [pharmacologic action]
- No associative relations
Cetirizine MeSH vs. SNOMED CT (1)

- Shared synonyms: MeSH 1/45, SNOMED CT 1/1
  - Cetirizine
- Shared relations: none
  - MeSH:
    - Isa: <chemistry>
    - Associative: <pharmacologic action>
  - SNOMED CT
    - Isa: <pharmacologic action>
    - Associative: none
Cetirizine  MeSH vs. SNOMED CT (2)
Summary

- Differing representations
  - Not necessarily inconsistent
  - Consistency may be difficult to assess automatically
- Often due to idiosyncratic representation in one ontology

- Hindrance to ontology alignment and evaluation methods relying on shared relations
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Extending Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

- Corpus terminology
- Identify terms in biomedical text (in reference to the UMLS)
- Identify additional terms
- Place these terms in UMLS hierarchies

[Bodenreider, ACL 2002]
Free cortisol in sepsis and septic shock.


Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@kuh.fi

BACKGROUND: Severe sepsis activates the hypothalamo-pituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according to the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. RESULTS: Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ($r = 0.90$, $P < 0.001$). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock (728 +/- 386 nmol/L vs 793 +/- 439 nmol/L, $P = 0.44$). Nonsurvivors had higher calculated serum free (209 +/- 151 nmol/L) and total (980 +/- 458 nmol/L) cortisol concentrations than survivors (119 +/- 111 nmol/L, $P = 0.002$, and 704 +/- 383 nmol/L, $P = 0.002$). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%. CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.
Identify UMLS concepts with MetaMap

http://skr.nlm.nih.gov/

Lister Hill National Center for Biomedical Communications
Interactive mode

Please Note: Users are responsible for compliance with the UMLS copyright restrictions.

To use this application, you must have signed the UMLS agreement. The UMLS agreement requires those who use the UMLS to file a brief report once a year to summarize their use of the UMLS. It also requires the acknowledgment that the UMLS contains copyrighted material and that those copyright restrictions be respected. The UMLS agreement requires users to agree to obtain agreements for EACH copyrighted source prior to its use within a commercial or production application.

[ Use of all the sources is permitted if the application is used for research purposes only. ]
Interactive MetaMap

The Interactive mode is only intended for the testing of the various programs and their options.
BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for
Select options

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Version (-V):</td>
<td></td>
</tr>
</tbody>
</table>

### Output Display Options
- Tagger Output (-T)
- Variants (-v)
- Plain Syntax (-p)
- Syntax (-x)
- Candidates (-c)
- Number Candidates (-n)
- Semantic Types (-s)
- Show CUIds (-l)
- Mappings (-m)
- Show Preferred Names Only (-O)
- MMI Output (-M)
- Machine Output (-q)
- Fielded Output (-f)
- Formal Tagger Output (-F)
- Fielded MMI output (-N)

### Output Display Options (continued)
- Show Original Phrases (-H)
- Show Concept's Sources (-G)
- Show Acronym/Abbreviations (-j)

### Behavior Options
- Tag Text (-t)
- No Derivational Variants (-d)
- Adj/Noun Derivalional Variants (-D)
- No Acronym/Abbreviation Variants (-a)
- Unique Acronym/Abbrev Variants (-u)
- Ignore Stop Phrases (-K) *(System Use)*
- Stop Large N (-l)
- Threshold (-r): [ ]
- Ignore Word Order (-i)

### Behavior Options (continued)
- Prefer Multiple Concepts (-Y)
- Best Mappings Only (-b)
- Truncate Candidates Mapping (-X)
- Use Word Sense Disambiguation (-y)

### Browse Mode Options
- Term Processing (-z)
- Allow Overmatches (-o)
- Allow Concept Gaps (-g)

### Misc. Options
- Display Phrases Only
- Dynamic Variant Generation (-8)
Run MetaMap

<table>
<thead>
<tr>
<th>&quot;Restrict to&quot; or &quot;Exclude&quot; Vocabulary Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Restrict to Sources (-R)</td>
</tr>
<tr>
<td>□ Exclude Sources (-e)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&quot;Restrict to&quot; or &quot;Exclude&quot; Semantic Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Restrict to Semantic Type(s) (-J)</td>
</tr>
<tr>
<td>□ Exclude Semantic Type(s) (-k)</td>
</tr>
</tbody>
</table>

[Submit Interactive MetaMap] [Reset Form]
Processing 00000000.tx.l: BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production.

Phrase: "Severe sepsis"

severe sepsis

Candidates

Meta Candidates (8):
1000 C1719672:Severe Sepsis [Disease or Syndrome]
861 C0036690:Sepsis (Septicemia) [Disease or Syndrome]
861 C0243026:Sepsis (Systemic infection) [Disease or Syndrome]
861 C1090821:Sepsis [Invertebrate]
789 C0333534:Septic [Functional Concept]
694 C0205082:Severe [Qualitative Concept]
694 C1519275:SEVERE (Severe Adverse Event) [Finding]
694 C1561581:Severe (Allergy Severity - Severe) [Finding]

Candidates

Mappings

Meta Mapping (1000):
1000 C1719672:Severe Sepsis [Disease or Syndrome]
Suggest term candidates

- Not recognized by MetaMap ay all
- Partially identified by MetaMap
- Missing terms in a concept
Suggest placement in UMLS

- Use a browser
- Identify close parent
- Examine its children
- Assess placement by comparing with potential siblings
Possible new terms (1)

◆ Hypothalamopituitary axis
  ● Concept exists: C0678897, but missing exact (neoclassical) synonym
    ■ hypothalamic pituitary axis
    ■ hypothalamus hypophysis axis
    ■ hypothalamus-pituitary axis

◆ American College of Chest Physicians
  ● Similar to other American Colleges (e.g., American College of Physicians ())
  ● Integrate as a child of Professional Organization or Group (C1522486 )
  ● NB: instance, cannot be a child of ACP
Possible new terms (2)

◆ Free cortisol
  ● Identified as a substance (C0443476), not a laboratory procedure / test result
    ■ Cortisol, free measurement (C0236401)

◆ Coolens method
  ● Missing term / concept
  ● Method for estimating (not measuring directly) the free fraction of cortisol
Possible new terms (3)

◆ Electrochemiluminescence immunoassay
  ● Missing concept
  ● Create as a child of Chemiluminescence assay (C0201709)

◆ Nonsurvivors
  ● Survivors exists as a concept (C0206194)
  ● Create as a child of Patients (C0030705)
<table>
<thead>
<tr>
<th>Day</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monday, June 9</strong></td>
<td>Introduction to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Biomedical Ontologies - Content and structure - Function</td>
</tr>
<tr>
<td><strong>Tuesday, June 10</strong></td>
<td>Interfaces to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Searching and Analyzing Biomedical Concepts</td>
</tr>
<tr>
<td></td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td><strong>Wednesday, June 11</strong></td>
<td>Critical Analysis of Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Extending Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Using Biomedical Ontologies
for Data Integration

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Motivation
◆ Some practical considerations and issues
  ● Integration approaches
  ● Concept repositories
  ● Using existing mappings
  ● Creating mappings through the UMLS
  ● Comparing semantic descriptions
◆ Thinking outside the integration box
Motivation
Motivation  Translational research

◆ “Bench to Bedside”
◆ Integration of clinical and research activities and results
◆ Supported by research programs
  ● NIH Roadmap
  ● Clinical and Translational Science Awards (CTSA)
◆ Requires the effective integration and exchange and of information between
  ● Basic research
  ● Clinical research
Translation research NIH Roadmap

Re-engineering the Clinical Research Enterprise

- Overview
- Implementation Group Members
- Funding Opportunities
- Funded Research
- Meetings
- Mid-course Reviews

TRANSLATIONAL RESEARCH

OVERVIEW

To improve human health, scientific discoveries must be translated into practical applications. Such discoveries typically begin at "the bench" with basic research — in which scientists study disease at a molecular or cellular level — then progress to the clinical level, or the patient's "bedside."

Scientists are increasingly aware that this bench-to-bedside approach to translational research is really a two-way street. Basic scientists provide clinicians with new tools for use in patients and for assessment of their impact, and clinical researchers make novel observations about the nature and progression of disease that often stimulate basic investigations.
Motivation  Translational research

Basic Research  Clinical Research and Practice

Lister Hill National Center for Biomedical Communications
Terminology and translational research

Cancer
Basic Research

NCI Thesaurus

SNOMED CT

EHR Cancer Patients

Lister Hill National Center for Biomedical Communications
Some practical considerations and issues

Integration approaches
Approaches to data integration

◆ Warehousing
  • Sources to be integrated are transformed into a common format and converted to a common vocabulary
  • Normalization through ontologies (e.g., GO annotations)

◆ Mediation
  • Local schema (of the sources)
  • Global schema (in reference to which the queries are made)
  • Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)
Some practical considerations and issues

*Concept repositories*
(Integrated) concept repositories

- Unified Medical Language System

- NCBO’s BioPortal

- Open Biomedical Ontologies (OBO)
  [http://obofoundry.org/](http://obofoundry.org/)

- caDSR
Integrating subdomains

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

OMIM

Biomedical literature

UMLS

MeSH

Model organisms

NCBI Taxonomy

GO

Genome annotations

Anatomy

FMA

Lister Hill National Center for Biomedical Communications
Trans-namespace integration

Addison's disease (363732003)
Clinical repositories

Other subdomains

SNOMED CT

Genetic knowledge bases

OMIM

Lister Hill National Center for Biomedical Communications

318

Genome annotations

GO

FMA

UMLS

C0001403

Biomedical literature

Addison Disease (D000224)

Anatomy

Model organisms

NCBI Taxonomy
Some practical considerations and issues

Mappings
Mappings

UMLS

NCI Thesaurus

SNOMED CT
Mappings

◆ Created manually
  ● UMLS

◆ Created automatically
  ● BioPortal

◆ Key to enabling semantic interoperability
◆ Enabling resource for the Semantic Web
Quality of mappings

- Created for a purpose
  - Reusability issues
- Generally unidirectional
  - Mapping from ontology 1 to ontology 2
  - Not necessarily reversible
Some practical considerations and issues

Comparing semantic descriptions
Semantic descriptions Consistent

S:Pancreatic structure \(\text{owl:sameAs}\) N:Pancreas

S:finding site \(\text{owl:sameAs}\) N:disease assoc. anat. site

S:Disorder of pancreas \(\text{owl:sameAs}\) N:Pancreatic disorder

S:Disorder of endocrine pancreas \(\text{owl:sameAs}\) N:Endocrine pancreas disorder

S:is_a \(\text{owl:sameAs}\) N:subClassOf

U:hasCUI

C0030286
SNOMED CT/NCI Thesaurus
Limited consistency

Lister Hill National Center for Biomedical Communications
Comparing formal definitions

- Relatively small proportion of relata in common between equivalent concepts from NCIt and SNOMED CT
- Large number of primitive concepts in NCIt and SNOMED CT (70-80%)
- Insufficient for effectively comparing definitions
  - Could not be used for validating the mapping provided by the UMLS

[Bodenreider, KRMED 2008]
Exercises
Exercise #1

◆ Check the equivalence (shared relata) between these 2 concepts:
  ● NCI Thesaurus: N:Endocrine pancreas disorder
  ● SNOMED CT: S:Disorder of endocrine pancreas
Exercise #2

◆ Find a correspondence in SNOMED CT for the LOINC term: *Sodium*:SCnc:-Pt:Ser/Plas:Qn

[the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

<table>
<thead>
<tr>
<th>Axis</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Sodium</td>
</tr>
<tr>
<td>Property</td>
<td>SCnc – Substance Concentration (per volume)</td>
</tr>
<tr>
<td>Timing</td>
<td>Pt – Point in time (Random)</td>
</tr>
<tr>
<td>System</td>
<td>Ser/Plas – Serum or Plasma</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn – Quantitative</td>
</tr>
<tr>
<td>Method</td>
<td>--</td>
</tr>
</tbody>
</table>
Comments on exercise #2

- Difficult in the absence of a search mechanism on the values of the relations
- Large number of underspecified descriptions in SNOMED CT
- 2 separate concepts for plasma and serum concentrations of sodium in SNOMED CT
- Property, time and scale not represented in SNOMED CT
Thinking outside the integration box

The Butte approach
Integrating genomic and clinical data

Genomic data

Clinical data

Upregulated genes

Diseases (extracted from text)

Coded discharge summaries

Laboratory data

Lister Hill National Center for Biomedical Communications
References

Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA