Unified Medical Language System

The graph behind the forest

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Biomedical trees
**Lineage** (full): root; cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus

- **Mus musculus** (house mouse) *Click on organism name to get more information.*

- **Mus musculus bactrianus** (southwestern Asian house mouse)
- **Mus musculus castaneus** (southeastern Asian house mouse)
- **Mus musculus domesticus** (western European house mouse)
- **Mus musculus gentilulus**
- **Mus musculus homourus**
- **Mus musculus molossinus** (Japanese wild mouse)
- **Mus musculus musculus** (eastern European house mouse)
- **Mus musculus praetextus**
- **Mus musculus wagneri**

Medical Subject Headings

- Amino Acids, Peptides, and Proteins [D12]
  - Proteins [D12.776]
    - Cytoskeletal Proteins [D12.776.220]
  - Contractile Proteins [D12.776.210]

- Amino Acids, Peptides, and Proteins [D12]
  - Proteins [D12.776]
    - Membrane Proteins [D12.776.543]

- Ankyrins [D12.776.543.080]
- Arrestins [D12.776.543.090] +
- Bacterial Outer Membrane Proteins [D12.776.543.100] +
- Calnexin [D12.776.543.162]
- Connexins [D12.776.543.225] +
- Dystrophin [D12.776.543.250]
- Dystrophin-Associated Proteins [D12.776.543.268] +
- Ephrins [D12.776.543.287] +
- Heterotrimeric GTP-Binding Proteins [D12.776.543.325] +

Gene Ontology

- all : all [189459]
  - GO:0008150 : biological_process [137743]
    - GO:0065007 : biological regulation [19935]
      - GO:0050789 : regulation of biological process [18154]
        - GO:0019222 : regulation of metabolic process [9816]
          - GO:0019216 : regulation of lipid metabolic process [108]
  - GO:0008152 : metabolic process [54684]
    - GO:0044238 : primary metabolic process [44818]
      - GO:0006629 : lipid metabolic process [3940]
        - GO:0019216 : regulation of lipid metabolic process [108]
  - GO:0019222 : regulation of metabolic process [9816]
    - GO:0019216 : regulation of lipid metabolic process [108]

http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
SNOMED Clinical Terms

Hierarchy for 'neoplasm of breast'

- neoplasm of thorax
- disorder of breast
  - neoplasm of breast
    + benign tumor of breast
    + carcinoma in situ of breast
    + malignant tumor of breast
    + neoplasm of female breast
    + neoplasm of male breast
    + neoplasm of skin of breast
    + neoplasm of uncertain behavior of breast

http://www.clininfo.co.uk/clue5/clue.htm
Biomedical trees revisited
Gene Ontology

- biological process
  - biological regulation
  - regulation of biological process
  - regulation of metabolic process
  - regulation of lipid metabolic process
  - metabolic process
    - primary metabolic process
    - lipid metabolic process

http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
SNOMED Clinical Terms

Disorder of Trunk

Disorder of Breast

Neoplasm of Thorax

Neoplasm of Breast

Neoplasm of Trunk

Hierarchies in SNOMED CT

http://www.clininfo.co.uk/clue5/clue.htm
Terminology integration

Unified Medical Language System
Addison’s disease in medical vocabularies

◆ Synonyms
  - Addisonian syndrome
  - Bronzed disease
  - Addison melanoderma
  - Asthenia pigmentosa
  - Primary adrenal deficiency
  - Primary adrenal insufficiency
  - Primary adrenocortical insufficiency
  - Chronic adrenocortical insufficiency

  eponym
  symptoms
  clinical
  variants
Organize terms

- **Synonymous terms clustered into a concept**
- **Preferred term**
- **Unique identifier (CUI)**

<table>
<thead>
<tr>
<th>Term</th>
<th>MeSH</th>
<th>D000224</th>
<th>MedDRA</th>
<th>10036696</th>
<th>ICD-10</th>
<th>E27.1</th>
<th>SNOMED CT</th>
<th>363732003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addison Disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary hypoadrenalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary adrenocortical insufficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addison's disease (disorder)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C0001403

Addison's disease
Diseases of the endocrine system

Diseases of the Adrenal Glands

Addison’s Disease
Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s Disease
Endocrine disorder

Adrenal disorder

Adrenal cortical disorder

Adrenal cortical hypofunction

Addison’s Disease
Endocrine disorder

Disorder of adrenal gland

Hypoadrenalism

Adrenal Hypofunction

Corticoadrenal insufficiency

Addison’s Disease
Organize concepts

- Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)
Adrenal Cortex Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Addison’s Disease

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Cortex Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Addison’s Disease

organize concepts

SNOMED
MeSH
AOD
Read Codes

UMLS
Endocrine Diseases

Adrenal Gland Diseases

Adrenal Cortex Diseases

Adrenal Cortex Dysfunction

Adrenal Gland Hypofunction

Hypoadrenalism

Secondary hypocortisolism

Adrenal cortical hypofunction

Addison’s Disease

Addison’s disease due to autoimmunity

Disorders of other endocrine gland

Other disorders of adrenal gland
Source Vocabularies

◆ 139 source vocabularies
  ● 17 languages
◆ Broad coverage of biomedicine
  ● 5.5M names
  ● 1.4M concepts
  ● 16M relations
◆ Common presentation
Biomedical forest
vs. graph
UMLS Knowledge Source Server

Quick Search
Select UMLS Release: 2007AA
Enter search value: Addison's disease

Advanced Searches
Metathesaurus Advanced Search
Facilitates advanced searching of the UMLS Metathesaurus, including restricting vocabularies, performing batch searches, performing XML queries, and using a command-line type interface.

http://umlsks.nlm.nih.gov/
Addison’s disease in UMLSKS (1)
Addison’s disease in UMLSKS (2)

<table>
<thead>
<tr>
<th>Ancestors:</th>
</tr>
</thead>
</table>

**Health and Disease**
- health and disease [G]
- disorder by body system or organ function [GK]
- endocrine disorder [GV]
- adrenal disorder [GV14]
  - adrenal cortical disorder [GV14.02]
  - adrenal cortical hypofunction [GV14.02.06]
  - Addison's disease [GV14.02.06.02]

**CRISP Thesaurus**
- disease/disorder [I]
  - endocrine disorder [I]
  - adrenal disorder [I]
  - hypoadrenalism [I]
  - Addison's disease [I]
- immunopathology [I]
  - autoimmune disorder [I]
  - Addison's disease [I]
  - Addison-Krankheit [I]
Addison’s disease in UMLSKS (3)

<table>
<thead>
<tr>
<th>Ancestors:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeS Me: Inc</td>
</tr>
<tr>
<td>CRI: dis en p a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alcohol and Other Drug Thesaurus</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-10</td>
</tr>
<tr>
<td>Endocrine, nutritional and metabolic diseases [I]</td>
</tr>
<tr>
<td>Disorders of other endocrine glands [I]</td>
</tr>
<tr>
<td>Other disorders of adrenal gland [I]</td>
</tr>
<tr>
<td>Primary adrenocortical insufficiency [I]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ICD-10 Austral Mod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endocrine, nutritional and metabolic diseases [I]</td>
</tr>
<tr>
<td>Disorders of other endocrine glands [I]</td>
</tr>
<tr>
<td>Other disorders of adrenal gland [I]</td>
</tr>
<tr>
<td>Primary adrenocortical insufficiency [I]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MedDRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endocrine disorders [I]</td>
</tr>
<tr>
<td>Adrenal gland disorders [I]</td>
</tr>
<tr>
<td>Adrenal cortical hypofunctions [I]</td>
</tr>
<tr>
<td>Addison's disease [I]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MedDRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolism and nutrition disorders [I]</td>
</tr>
<tr>
<td>Metabolism disorders NEC [I]</td>
</tr>
<tr>
<td>Metabolic disorders NEC [I]</td>
</tr>
<tr>
<td>Addison's disease [I]</td>
</tr>
</tbody>
</table>
Addison’s disease in UMLSKS (4)

<table>
<thead>
<tr>
<th>Ancestors:</th>
<th>MeSH Me: Inc D</th>
<th>Alcohol and Other Drug Theorems</th>
<th>ICD-10</th>
</tr>
</thead>
</table>
Addison’s disease in UMLSKS (5)

SNOMED Clinical Terms
- Clinical finding
  - Finding by site
    - Finding of body region
    - Finding of trunk structure
  - General finding of abdomen
  - Abdominal organ finding
  - Disorder of adrenal gland
    - Disorder of adrenal cortex
  - Adrenal cortical hypofunction
    - Primary adrenocortical insufficiency
UMLS Semantic Navigator

Siblings

Concepts & Ideas
- Clinical Syndromes

Disorders
- Aarskog syndrome
- Abstinence Syndrome
- Achard Syndrome
- Acroparesthesia Syndrome
- ACTH deficiency
- Addisonian crisis
- Adie Syndrome
- Adrenal cortex atrophy
- Adrenal insufficiency due to adrenal

AmiGO

- GO:0008150 : biological_process [137743]
- GO:0065007 : biological regulation [19935]
- GO:0050789 : regulation of biological process [18154]
- GO:0019222 : regulation of metabolic process [9816]
- GO:0019216 : regulation of lipid metabolic process [108]

- GO:0008152 : metabolic process [54684]
- GO:0044238 : primary metabolic process [44818]
- GO:0006629 : lipid metabolic process [3940]
- GO:0019216 : regulation of lipid metabolic process [108]

- GO:0019222 : regulation of metabolic process [9816]
- GO:0019216 : regulation of lipid metabolic process [108]

http://amigo.geneontology.org/cgi-bin/amigo/go.cgi
Semantics of the UMLS graph

Issues and challenges
Visualization of large graphs

disorders NEC
Infectious disease of abdomen
Endocrine infection
Tuberculosis of other organs NOS
Non-Neoplastic Adrenal Gland Disorder

inflammation of specific body structures or tissue

Tuberculosis of adrenal glands

Syndrome
Adrenal gland hypofunction
Autoimmune

Addison's disease

Tuberculotic Addison's disease
Acyclicity

“back edge” from a child concept to a parent concept

Reflexive 13,000

Direct 1800

Indirect 120
Underspecification of relationships

- Relationship “attribute” not always present
- Relations used to create hierarchies vs. hierarchical relations

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] +
Which tasks?

- Information integration
- Mapping

- Depending on the degree of human involvement
  - Hypothesis generation / validation
  - Knowledge discovery
  - Automated reasoning

- Knowledge standardization
  - Common format
  - Common semantics
Which formalisms?

- **SKOS – Thesaurus**
  - Simple Knowledge Organization Schema
- **RDF – Concept-Relationship-Concept triples**
  - Resource Description Framework
- **Description Logics / Frames**
  - OWL  Web Ontology Language
  - Protégé (frames / OWL)
  - OBO  Open Biomedical Ontology
- **Rule languages**
- **Formal logic**
Which identifiers?

◆ For concepts
  ● Namespaces, ontologies, knowledge bases
    ■ OBO – Open Biomedical Ontologies
    ■ UMLS – Unified Medical Language System
    ■ NCBI Entrez (Entrez Gene, GenBank, UniGene, …)
  ● Mappings across information sources

◆ For relationships
Conclusions
Integrating subdomains

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

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Other subdomains

Biomedical literature

Genome annotations

Anatomy

Model organisms

Other subdomains

Clinical repositories

Genetic knowledge bases
From *glycosyltransferase* to *congenital muscular dystrophy*
UMLS References

- **UMLS**
  umlsinfo.nlm.nih.gov

- **UMLS browsers**
  (free, but UMLS license required)
  - Semantic Navigator:
  - RRF browser
    (standalone application distributed with the UMLS)
Gentle introduction


Seminal paper

Biomedical information integration through RDF

◆ Biomedical perspective

◆ Semantic Web perspective