The Unified Medical Language System: Between Terminology and Ontology
Medical Ontology Research
Terminology  Adrenal gland diseases

- Adrenal gland diseases: MeSH D000307
- Adrenal disorder: AOD 0000005418
- Disorder of adrenal gland: Read C15z.
- Diseases of the adrenal glands: SNOMED DB-70000
Ontology Adrenal gland diseases

Adrenal Gland Disease

- Adrenal Abnormality
- Abdominal Disorder
- Endocrine Disorder

Phenomenon

GALEN.opengalen.org

Adrenal Gland Disease involves Suprarenal Gland

has status pathological

Addison’s Disease Adrenal Tumor Congenital Adrenal Hyperplasia
<table>
<thead>
<tr>
<th>Adrenal gland diseases</th>
<th>MeSH</th>
<th>D000307</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenal disorder</td>
<td>AOD</td>
<td>0000005418</td>
</tr>
<tr>
<td>Disorder of adrenal gland</td>
<td>Read</td>
<td>C15z.</td>
</tr>
<tr>
<td>Diseases of the adrenal glands</td>
<td>SNOMED</td>
<td>DB-70000</td>
</tr>
</tbody>
</table>
Biomedical knowledge organization

Terminologies

Semantic Spaces

Ontologies

Medical Subject Headings
International Classification of Diseases
SNOMED

UMLS

Cyc, WordNet
GALEN
Digital Anatomist

Lister Hill National Center for Biomedical Communications
Outline

- Overview
- Benefits
- Limitations
Overview
Biomedical terminologies

◆ Core vocabularies
  ● anatomy (UWDA, Neuronames)
  ● drugs (First DataBank, Micromedex)
  ● medical devices (UMD, SPN)

◆ Several perspectives
  ● clinical terms (SNOMED, CTV3)
  ● information sciences (MeSH, CRISP)
  ● administrative terminologies (ICD-9-CM, CPT-4)
  ● standards (HL7, LOINC)
Biomedical terminologies (cont’d)

- Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
- Knowledge bases (AI/Rheum, DXplain, QMR)
UMLS

- Two-level structure
  - Semantic Network
    - 135 Semantic Types (STs)
    - 54 types of relationships among STs
  - Metathesaurus
    - 875,000 concepts
    - ~12 M inter-concept relationships
  - Link = categorization
- Lexical resources
UMLS Services

- Lexical tools (e.g., normalization)
- Browsers
- MetaMap
- API (Java, XML)
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
AD in UMLS  Contexts

- Endocrine Diseases
  - Adrenal Gland Diseases
    - Adrenal Cortex Diseases
      - Hypoadrenalism
        - Adrenal Gland Hypofunction
          - Adrenal cortical hypofunction
            - Addison’s Disease
AD in UMLS  SNOMED context

Endocrine Diseases

Adrenal Gland Diseases

Addison’s Disease
AD in UMLS  MeSH context

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s Disease
AD in UMLS  Read Codes context

- Endocrine Diseases
  - Adrenal Gland Diseases
    - Hypoadrenalism
      - Adrenal Gland Hypofunction
        - Adrenal cortical hypofunction
          - Addison’s Disease
AD in UMLS  AOD Thes. context

- Endocrine Diseases
  - Adrenal Gland Diseases
    - Adrenal Cortex Diseases
      - Adrenal cortical hypofunction
        - Addison’s Disease
Benefits
UMLS compared to individual vocabularies

- Broader scope
- Extended coverage
- Finer granularity
- Unique identifier
- Synonymous terms clustered into concepts
- Additional synonyms
- Additional hierarchical relationships
- Semantic categorization
Direct benefits

- Concept categorization
- Information retrieval
  - Synonyms
  - Cross-language features
- Information extraction
  - MetaMap
  - Normalization
- Information visualization
  - Knowledge Source Server
  - Semantic Navigator
UMLS Semantic Navigator
Siblings

Disorders

- Acquired Immuno deficiency Syndrome
- Acute adrenal insufficiency
- Addisonian crisis
- Adrenal Gland Hyperfunction
- Adrenal insufficiency due to adrenal metastasis
- Allergic/autoimmune thyroiditis
- Allergic arthritis
- Angelman Syndrome
- Antiphospholipid Syndrome
- Aortic Diastole
- Autoerythrocyte sensitivity disorder, HUS
- Autonomic Diseases of the Nervous System
- Autoimmune hemolytic anemia
- Autoimmune leukopenia
- Autoimmune pancytopenia
- Autoimmune Thrombocytopenia
- Battered Child Syndrome
- Behcet Syndrome
- Bloom Syndrome
- Bitter Adipose
- Cerebral Adipose
- Capsaicin
- Chronic Six
- Congenital hypoplasia of adrenal gland
- Congenital hypoplasia of adrenal gland
- CRYOMEGALOCOLIC SYNDROME
UMLS Semantic Navigator Guru concepts

Other Related Concepts

Disorders

- Addisonian crisis
- addisons Disease
- secondary to Adrenal destruction
- addisons Disease
- secondary to idiopathic atrophy
- adrenal cortical hypofunction
- autoimmune syndrome type II, polyglandular
- endo/metab problem
- hypoglycemia
- hyponatremia
- tuberculosis
- tuberculosis of adrenal glands
- tuberculous Addison's
UMLS Semantic Navigator  Concepts

Co-occurring Concepts

Anatomy

- Adrenal Cortex [12]
- Adrenal Glands [19]
- Ear Cartilages [2]
- Ear, External [2]
- Liver [2]
- Pituitary Gland [3]
- Tears body
  substance [2]
- X Chromosome [3]

Chemicals & Drugs

- Alanine
- Transaminase [2]
- Aldosterone [3]
- Anti-Inflammatory Agents, Steroidal [2]
- Antigens,
UMLS Semantic Navigator

Relationships
of Addison's Disease (C1)
Disease or Syndrome
to Adrenal Cortex (C2)
Body Part, Organ, or Organ Component

Metathesaurus Relationships
C1 co-occurs with C2
Frequency = 12

Semantic Network Relationships
Disease or Syndrome
• has_location
Body Part, Organ, or Organ Component

Close this window

Interface version: 2.01   UMLS data: UMLS_2000
Additional (indirect) benefits

- **Examples**
  - Mapping across vocabularies
  - Semantics of statistical associations
  - Redundancy in hierarchical relations
Additional (indirect) benefits

Mapping across vocabularies
Terminology integration

- Terminology integration is a step towards interoperability
  - Clusters of synonyms from different sources
  - Paths between terms from different sources
Indexing Initiative

- For noun phrases extracted from medical texts, map to UMLS concepts
- Then, select from the MeSH vocabulary the concepts that are the most closely related to the original concepts

Medical text ➔ Noun phrase ➔ MeSH descriptor ➔ UMLS

[Aronson & al., AMIA, 2000]
Restrict to MeSH

- Based on the principle of semantic locality
- Use different components of the UMLS
- 4 techniques of increasing aggressiveness
  - Use Synonymy  MRCON + MRSO
  - Use Associated expressions (ATXs)  MRATX
  - Explore the Ancestors  MRREL + SN
  - Explore the Other related concepts  MRREL + SN

[Bodenreider & al., AMIA, 1998]
Restrict to MeSH: Synonymy

- Term mapped to Source concept
- For this concept, is there a synonym term that comes from MeSH? (MRSO)
Restrict to MeSH: Assoc. expressions

- If not,
- Is there an associated expression (ATX) that describes this concept using a combination of MeSH descriptors? *(MRATX)*

Endoscopic removal of intraluminal foreign body from oesophagus without incision

Diagram:
- AND
- MH/SH
  - Esophagus
  - surgery
  - Foreign Bodies
Restrict to MeSH: Ancestors

- If not, let us build the graph of the ancestors of this concept
  - using parents and broader concepts (MRREL)
  - all the way to the top
  - excluding ancestors whose semantic types are not compatible with those of the source concept (MRSTY)
- From the graph, select the concepts that come from MeSH (MRSO)
- Remove those that are ancestors of another concept coming from MeSH
Restrict to MeSH: Other related concepts

◆ If not, explore the other related concepts (MRREL) whose semantic types are compatible with those of the source concept (MRSTY)
◆ From those, select the concepts that come from MeSH (MRSO)
Restrict to MeSH: Example

Vein of neck, NOS

There is a MeSH term in the synonyms of SC

SC is described by a combination of MeSH terms (ATX)

The ancestors of SC contain MeSH terms

MeSH terms from non-hierarchically related concepts

Vein + Neck
Restrict to MeSH: Example

Head and neck, NOS
Body part, NOS
Blood Vessels
Vascular structure

Neck
Head

Veins
Systemic veins

Vein of head and neck, NOS
Vein of neck, NOS
Additional (indirect) benefits

Semantics of statistical associations
Co-occurrence Overview

- Co-occurrence between MeSH descriptors in MEDLINE citations
- 8 M pairs of co-occurring concepts
- Implicit semantics
- The UMLS provides knowledge for helping make this relationship explicit
  - Corresponding symbolic knowledge (Metathesaurus)
  - Categorization (Semantic Network)

[Burgun & al., MEDINFO, 2001]
Co-occurrence Example

Adrenal cortical hypofunction

Addison's disease

Co-occurrence (frequency = 20)

Cortisol

Adrenal gland

location of

cortex

produces

medulla
Co-occurrence Methods

◆ Based on Metathesaurus relationships
  ● Does “Cortisol” belong to the family of “Addison’s disease”?

◆ Based on Semantic Network relationships
  ● What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?

Addison's disease

Co-occurrence (frequency = 20)

Cortisol
What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?

Does “Cortisol” belong to the family of “Addison’s disease”?

Does “Cortisol” belong to the family of “Addison’s disease”??
Redundancy in hierarchical relations
AD in UMLS  Redundancy

- Adrenal Gland Hypofunction
- Adrenal cortical hypofunction
- Addison’s Disease
AD in UMLS  Redundancy

- Adrenal Gland Hypofunction
- Adrenal cortical hypofunction
- Addison’s Disease

Read, UMLS

AOD, Read, UMLS ICD-9-CM

MeSH UMLS
Redundancy index

\[ \text{IR}_{AD} = \text{IRP}_{AD} + \text{IRP}_{ACD} + \text{IRP}_{ABCD} = 4 \]

with:
\[
\begin{align*}
\text{IRP}_{AD} &= \text{NS}_{AD} = 1 \\
\text{IRP}_{ACD} &= \min(\text{NS}_{AC}, \text{NS}_{CD}) = \min(1, 2) = 1 \\
\text{IRP}_{ABCD} &= \min(\text{NS}_{AB}, \text{NS}_{BC}, \text{NS}_{CD}) = \min(2, 3, 2) = 2
\end{align*}
\]

[Bodenreider & al., AMIA, 2003 (subm.)]
Limitations
Limitations

- Structural inconsistency
  - Cycles in the graph of hierarchical relations

- Semantic inconsistency
  - Between Metathesaurus and Semantic Network
  - Meaning of *isa*

- Missing relations
  - Synonymy
  - Hierarchical relations (missing or underspecified)

[Cimino, *JAMIA*, 1998]
Compensation mechanisms

◆ Examples
  ● Removing cycles from hierarchical relations
  ● Lexically-suggested hyponymic relations
Compensation mechanisms

Removing cycles
from hierarchical relations
Hierarchies in source vocabularies

- Often task-driven rather than based on principles
- Usually suitable for information retrieval
  - Better recall
  - Precision may not be crucial
- Not necessarily suitable for reasoning
- But expected to be consistent structurally
AD in UMLS

- Multiple tree structures combined into a graph structure
- Directed acyclic graph (DAG)
Actually, there are some cycles

- Anti-infective Agents
  - Disinfectants and Cleansers
    - Disinfectants
      - Disinfectant soap
        - Germicidal soap
Issues with cycles

◆ Theoretical
  ● Violate the antisymmetry property of partial ordering relations

◆ Practical
  ● Loops in graph traversal
  ● Impossible to perform transitive reduction
Cycle due to underspecification

- Specified and underspecified terms
  - May appear at different levels in a source hierarchy
  - Are clustered into the same concept (same meaning)
Other causes

- Compound terms
- Metadata
  - HYDROCELE, Hydrocele
- Classes and member
  - Purines, Purine
- Organizational conventions
  - Acid + Base ↔ Salt + Water
- Idiopathic
  - Wrong relationships
  - Use of non-hierarchical relationships in “hierarchies”

[Bodenreider, AMIA 2001]
Compensation mechanisms

Lexically-suggested hyponymic relations
Methods Overview

- Syntactic analysis to identify adjectival modifiers
- Generate transformed terms by removing adjectival modifiers
- Map transformed terms to the UMLS
- Study the relationship between original term and transformed term in the UMLS, if any

[Bodenreider & al., T/A, 2001]
Identify adjectival modifiers

- Underspecified syntactic analysis
  - Xerox part of speech tagger
  - SPECIALIST Lexicon (UMLS)
- Modifiers used: adjectives (+ adverbs)
- Modifiers identified in 64% of the terms
- Usually 1 to 2 modifiers
- Unique modifiers
  - 5400 adjectives
  - 69 adverbs
Transforming terms

- Remove any combination of modifiers found in the original term
- $2^n - 1$ transformed terms when the original term has $n$ modifiers
- 104,000 transformed terms generated
Mapping transformed terms to UMLS

- Increasing aggressiveness
  - Exact match
  - After normalization
- 25% of the transformed terms successfully mapped to UMLS
Excluding non-hyponymic relations

- If in hyponymic relation, original term and the transformed term should have the same semantic type (both Disease or both Procedure)
- Different semantic types in 10%

Diagram:
- palate
- cleft palate
- Body part
- Disorder
Checking relationships against UMLS

- Original term (OT) – Transformed term (TT)
  - Synonyms (same concept)
  - TT ancestor of OT
  - Siblings (inter-concept relationship)
  - Otherwise related
Lexically-suggested relationships / UMLS

- 28,851 pairs of terms
  - Original SNOMED term
  - Transformed term (found in UMLS)

- Corresponding relationship in the Metathesaurus
  - Hierarchical in 50% of the cases
  - « Sibling » in 25% of the cases
  - Missing in 25% of the cases
More limitations

- Some missing / wrong relations are hard to detect
- Some relations are present but hard to find
Not all “isa” relationships are transitive!

Autoimmune Diseases

Addison’s disease

Tuberculous Addison’s disease

Addison’s disease due to autoimmunity

is generally a
Semantic Network

- Disease or Syndrome
  - Somatic Disease
  - Mental Disease
  - Somatic and Mental Disease

No hybrid type
Thus, multiple categorization

[Burgun & al., FOIS, 2001]
Semantic Network

Disease or Syndrome

Somatic Disease
Mental Disease

Vascular Dementia

Ad hoc precision in hierarchies
Thus, no “Somatic disease” type
Semantic Network

Disease or Syndrome

“Other” Disease

Mental Disease

Vascular Dementia

No “other” type;
Assign to the common supertype
Semantic Network

Disease or Syndrome

Mental Disease

Diabetes Mellitus

Vascular Dementia
Possible solutions

◆ Description logics
◆ Natural Language Processing
  (semantic interpretation of the terms)
◆ Comparing knowledge sources
  (alignment, inference)
Conclusions
Conclusions  The up side

- Terminology integration is a step towards interoperability
  - Clusters of synonyms from different sources
  - Paths between terms from different sources
Conclusions The down side

- However, interoperability requires more than loosely aligned terminologies
- The UMLS does not claim to be an ontology
- The UMLS is, however, a resource for acquiring biomedical ontologies
Conclusions

Medical Ontology Research

Terminologies

Semantic Spaces

UMLS

Ontologies

information retrieval

reasoning
Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: etbsun2.nlm.nih.gov:8000

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