Biomedical knowledge organization

Terminologies

Semantic Spaces

UMLS

Medical Subject Headings
International Classification of Diseases
SNOMED

[...]

Ontologies

Cyc
WordNet
Digital Anatomist
[...]

Lister Hill National Center for Biomedical Communications
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)
Unified Medical Language System
UMLS Metathesaurus

◆ 50 families of vocabularies
◆ 800,000 concepts
◆ 1.7 M names
◆ About 10 M inter-concept relationships
  ● Symbolic
  ● Statistical
Naming and Organizing
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

- Different names
- Different contexts
AD in medical vocabularies

Names

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

  eponym
  symptoms
  clinical
  variants
AD in UMLS  Names

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

<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>
<tr>
<td></td>
<td></td>
<td>C0001621</td>
</tr>
</tbody>
</table>

C0001621

Adrenal Gland Diseases
AD in medical vocabularies

Contexts

SNOMED

Diseases

Endocrine Diseases

Adrenal Gland Diseases

Addison’s Disease

MeSH

Diseases

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Immunologic Diseases

Autoimmune Diseases

Addison’s Disease
AD in UMLS  Contexts

- Multiple tree structures combined into a graph structure
- Directed acyclic graph (DAG)
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
Biomedical terms  Complexity

◆ Mostly precoordinated
  Autoimmune disorders
  Addison’s disease
  Addison’s disease due to autoimmunity

◆ Various complexity
  Motor vehicle traffic accident involving collision, on highway, between
  motor vehicle and fallen stone, traffic sign, tree or utility pole,
  passenger of motor vehicle injured

◆ Neoclassical compounds
  Open prostatectomy
Biomedical terms  Variability

◆ Case
◆ Punctuation
◆ Underspecification feature
◆ Genitive mark
◆ Inflection
◆ Word order
Normalization Example

Hodgkin Disease
HODGKINS DISEASE
Hodgkin's Disease
Disease, Hodgkin's
Hodgkin's, disease
HODGKIN'S DISEASE
Hodgkin's disease
Hodgkins Disease
Hodgkin's disease NOS
Hodgkin's disease, NOS
Disease, Hodgkins
Diseases, Hodgkins
Hodgkins Diseases
Hodgkins disease
hodgkin's disease
Disease, Hodgkin

normalize
disease hodgkin
Normalization

- Remove genitive: Hodgkin’s diseases, NOS
- Remove stop words: Hodgkin diseases, NOS
- Lowercase: Hodgkin diseases,
- Strip punctuation: hodgkin diseases,
- Uninflect: hodgkin diseases
- Sort words: hodgkin disease
- hodgkin disease
- disease hodgkin
Lexical knowledge

Adrenal gland diseases
- Adrenal disorder
- Disorder of adrenal gland
- Diseases of the adrenal glands
  C0001621

Diseases
Endocrine Diseases

Adrenal cortical hypofunction
Addison’s disease
- Primary adrenocortical insufficiency
  C0001403

Addison’s Disease
Adrenocortical insufficiency
  C0405580

Endocrine Diseases

Lister Hill National Center for Biomedical Communications
Lexical resources

◆ Lexicon

| Syntactic Category: noun
| Inflection Type: reg

Base Form: gland

Singular: gland
Plural: glands

◆ Lexical tools

● stop words
Diseases of the adrenal glands

● word order
Diseases of the adrenal glands
Adrenal glands diseases

● inflection
gland → glands

● derivation
cortex → cortical
Synonymy
From terms to concepts

- Terms are organized by meaning
- Clusters of synonymous terms
  - UMLS (concepts)
  - WordNet (synsets)
- Concepts are defined by extension
- Polysemous terms may belong to several concepts
Synonymy and Granularity

- **Splitters vs. Lumpers**
  - Formal vs. practical definition of synonymy

- **Organization vs. Representation**
  - Task-oriented vs. Purpose-independent
  - Generic vs. prototypical

- Posttransfusion hepatitis
  - Posttransfusion **viral** hepatitis

- Asplenia
  - **Congenital** asplenia
Implicit contextual knowledge
Disambiguate according to the context
Hierarchical Relationships
Hierarchy

◆ Hierarchical relationships
  ● Taxonomy (isa)
  ● Meronomy (part of)
◆ Partial ordering
  ● [Reflexivity]
  ● Antisymmetry
  ● Transitivity
◆ Inheritance
◆ Reasoning
Principles of subsumption

Aneuvrysm

isa

Aortic Aneuvrysm

isa

Aortic Aneuvrysm, Thoracic

isa

Thoracoabdominal Aortic Aneuvrysm

aneuvrysm

an. of the aorta

partitive refinement of a concept element

introduction of a specializing criterion

an. of the thoracic aorta

conjunctive coordination

an. of the thoracic aorta and abdominal aorta
Hierarchies in source vocabularies

◆ Structure
  ● Single tree
  ● Polyhierarchical (multiple parents)

◆ Relationships
  ● Usually implicit
  ● May be other than *isa* or *part of*
    ■ E.g., Thesaurus relationships
Hierarchies in source vocabularies

- Often task-driven rather than based on principles
- Usually suitable for information retrieval
  - Better recall
  - Precision may not be crucial
- Generally not suitable for reasoning
Some “isa” relationships are symmetric!

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

**Diagram:**
- Fever
  - Fever, unspecified
  - Fever of unknown origin
- Fever of unknown origin
- ICD-10
- MeSH
Not all “isa” relationships are transitive!

- Autoimmune Diseases
- Addison’s disease
- Tuberculous Addison’s disease
- Addison’s disease due to autoimmunity

*is generally a*
Mapping relationships
(Cross-reference)
Internal mapping relationships

- Used to implement multiple inheritance in monoaxial systems

- **Joint disease**
  - Arthritis
  - Tuberculosis
  - Tuberculous arthritis

- **Infectious disease**
  - Arthritis
  - Tuberculosis
  - Tuberculous arthritis (*)
  - Tuberculous arthritis (†)

- **Joint disease**
  - Arthritis
  - Tuberculosis
  - Tuberculous arthritis (*)
  - Tuberculous arthritis (†)
External mapping relationships

- Created to translate between medical vocabularies
  - SNOMED to ICD-9-CM
  - MeSH Supp. concepts to Descriptors
  - ICD-9-CM to MeSH

- Weak semantics
  - IR relationships
  - Clinical relationships (“clinically associated with”)

Lister Hill National Center for Biomedical Communications
Alternative approaches
Concepts

◆ Post-coordination
  ● atomic concepts
  ● compositionality

◆ Model-oriented terminologies
  ● conceptual structures
  ● description logics
Terminologies

◆ GALEN
  - EU project (Rector & al.)
  - Strong emphasis on the model
  - Limited coverage
  - Support for terminology services

◆ SNOMED-RT/CT
  - Developed by CAP
  - Based on SNOMED
  - Additional structure (reference terminology)
  - Collaboration with CTV3 (formerly Read Codes)
However…

- Less intuitive
- Steep learning curve
- Not widely deployed yet
- Very expensive to build
- Less comprehensive
Conclusions
Conclusion

◆ Biomedical domain is vast
  ● Broad and deep
  ● 800,000 UMLS concepts
◆ Difficult to organize consistently
◆ Organization of knowledge needs to be principled to support reasoning
◆ Evolve
  ● From structures supporting IR
  ● To structures supporting reasoning
References

Bodenreider O., Bean C. A.  
Relationships among knowledge structures: Vocabulary integration within a subject domain.  
Contact: olivier@nlm.nih.gov

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