The Unified Medical Language System

*What is it and how to use it?*

Olivier Bodenreider, MD, PhD

Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Part I: *What is the UMLS?*
- Introduction
- Overview through an example
- The three UMLS Knowledge Sources
  - UMLS Metathesaurus
  - UMLS Semantic Network
  - SPECIALIST Lexicon and lexical tools

Part II: *How to use the UMLS?*
- A UMLS-based algorithm
Part I
What is the UMLS?

(1) Introduction
What does UMLS stand for?

- Unified
- Medical
- Language
- System

UMLS®
Unified Medical Language System®
UMLSS Metathesaurus®
Motivation

- Started in 1986
- National Library of Medicine
- “Long-term R&D project”
- Complementary to IAIMS (Integrated Academic Information Management Systems)

«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.
- The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.
- The second is the distribution of useful information among many disparate databases and systems.»
The UMLS in practice

- **Database**
  - Series of relational files

- **Interfaces**
  - Web interface: Knowledge Source Server (UMLSKS)
  - Application programming interfaces (Java and XML-based)

- **Applications**
  - lvg (lexical programs)
  - MetamorphoSys (installation and customization)
  - RRF browser (browsing subsets)

**The UMLS is not an end-user application**
Part I
What is the UMLS?

(2) Overview through an example
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
  • Melasma addisonii
  • Asthenia pigmentosa
  • Primary adrenal deficiency
  • Primary adrenal insufficiency
  • Primary adrenocortical insufficiency
  • Chronic adrenocortical insufficiency

◆ Contexts: different hierarchies
  | eponym
  | symptoms
  | clinical
  | variants

NLM
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
Endocrine system diseases

Adrenal gland diseases

Adrenal Insufficiency

Addison Disease
Endocrine disorders

Adrenal gland disorders

Adrenal cortical hypofunctions

Addison's Disease
Disorder of endocrine system

Disorder of adrenal gland

- Hypoadrenalism
- Adrenal hypofunction
- Adrenal cortical hypofunction

- Disorder of adrenal cortex

Addison's Disease
Disorder of endocrine system

Disorder of adrenal gland

Adrenal gland hypofunction

Disorder of adrenal cortex

Adrenal cortical hypofunction

Addison's Disease
NCI Thesaurus

- Non-neoplastic endocrine disorder
  - Adrenal gland disorder
    - Adrenal gland insufficiency
      - Adrenal cortical insufficiency
        - Addison's Disease
ICD-10

- Endocrine, nutritional and metabolic diseases
  - Disorders of other endocrine glands
    - Other disorders of adrenal gland
      - Primary adrenocortical insufficiency
Organize concepts

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

Disease

Endocrine system diseases

Adrenal gland diseases

Adrenal gland hypofunction

Adrenal cortical hypofunction

Addison's Disease

Adrenal cortex diseases

SNOMED CT
SNOMED Intl
MeSH
MedDRA
- Endocrine system diseases
  - Non-neoplastic endocrine disorder
    - Non-neoplastic adrenal gland disorder
      - Adrenal gland hypofunction
      - Adrenal cortical hypofunction
        - Addison's Disease
  - Adrenal gland diseases
    - Adrenal cortex diseases
  - Other disorders of adrenal gland
  - Disorders of other endocrine glands
  - Endocrine / nutritional / metabolic disorder
Relate to other concepts

◆ Additional hierarchical relationships
  - link to other trees
  - make relationships explicit
◆ Non-hierarchical relationships
◆ Co-occurring concepts
◆ Mapping relationships
Categorize concepts

- High-level categories (semantic types)
- Assigned by the Metathesaurus editors
- Independently of the hierarchies in which these concepts are located

Diagram:

```
Disease or Syndrome
   /\   
  /   \   
Endocrine Diseases
 /     \     
Adrenal Gland Diseases
 /         \         
Adrenal Gland Hypofunction
 /             \             
Addison’s Disease
```
How do they do that?

- Lexical knowledge
- Semantic pre-processing
- UMLS editors
Lexical knowledge

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

- Metadata in the source vocabularies
- Tentative categorization
- Positive (or negative) evidence for tentative synonymy relations based on lexical features
Additional knowledge: UMLS editors

Adrenal gland diseases
  └── Adrenal gland hypofunction
      └── Adrenal cortical hypofunction
          └── Addison's Disease
  └── Adrenal cortex diseases
  └── Other disorders of adrenal gland
UMLS Summary

- Synonymous terms clustered into concepts
- Unique identifier
- Finer granularity
- Broader scope
- Additional hierarchical relationships
- Semantic categorization
Part I
What is the UMLS?

(3) UMLS Metathesaurus
Unified Medical Language System

- **SPECIALIST Lexicon**
  - 360,000 lexical items
  - Part of speech and variant information

- **Metathesaurus**
  - 6M names from over 100 terminologies
  - 1.5M concepts
  - 8M relations

- **Semantic Network**
  - 135 high-level categories
  - 7000 relations among them
Metathesaurus Basic organization

◆ Concepts
  ● Synonymous terms are clustered into a concept
  ● Properties are attached to concepts, e.g.,
    ■ Unique identifier
    ■ Definition

◆ Relations
  ● Concepts are related to other concepts
  ● Properties are attached to relations, e.g.,
    ■ Type of relationship
    ■ Source
Source Vocabularies

- 141 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 6.1M names
  - 1.5M concepts
  - 8M relations
- Common presentation
Biomedical terminologies

- **General vocabularies**
  - anatomy (UWDA, Neuronames)
  - drugs (RxNorm, First DataBank, Micromedex)
  - medical devices (UMD, SPN)

- **Several perspectives**
  - clinical terms (SNOMED CT)
  - information sciences (MeSH, CRISP)
  - administrative terminologies (ICD-9-CM, CPT-4)
  - data exchange terminologies (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)

- Terminology of knowledge bases (AI/Rheum, DXplain, QMR)

The UMLS serves as a vehicle for the regulatory standards (HIPAA, CHI)
Integrating subdomains

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

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

- Clinical repositories
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Genetic knowledge bases
  - NCBI Taxonomy
  - Other subdomains

- Biomedical literature
  - OMIM
  - MeSH

- Model organisms
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Genome annotations
  - NCBI Taxonomy
  - Other subdomains

- Addison's disease (363732003)

- Other subdomains
  - NCBI Taxonomy
  - Other subdomains

- Model organisms
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Genome annotations
  - NCBI Taxonomy
  - Other subdomains

- Clinical repositories
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Biomedical literature
  - OMIM
  - MeSH

- Model organisms
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Genome annotations
  - NCBI Taxonomy
  - Other subdomains

- Clinical repositories
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Biomedical literature
  - OMIM
  - MeSH

- Model organisms
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Genome annotations
  - NCBI Taxonomy
  - Other subdomains

- Clinical repositories
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Biomedical literature
  - OMIM
  - MeSH

- Model organisms
  - SNOMED CT
  - OMIM
  - UMLS
  - FMA
  - GO
  - Anatomy

- Genome annotations
  - NCBI Taxonomy
  - Other subdomains
Addison’s Disease

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.

Disease or Syndrome

Addison’s Disease

ADRENAL INSUFFICIENCY (ADDISON’S DISEASE)
ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE
Hypoadrenalisms, Primary
Melasma addisonii
Primary adrenal deficiency
Asthenia pigmentosa
Bronzed disease
Insufficiency, adrenal primary
Primary adrenocortical insufficiency
Addison’s, disease

Maladie d’Addison - French
Addison-Krankheit - German
Morbo di Addison - Italian
Doença de Addison - Portuguese
АДДИСОНОВА БОЛЕЗНЬ - Russian
アジソン病 - Japanese

Speaking of Addison’s Disease:

SNOMED CT
SNOMED Intl
MeSH
MedDRA

C0001403
Part II
How to use the UMLS?

A *UMLS*-based algorithm
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
Restrict to MeSH

- Based on the principle of *semantic locality*
- Use different components of the UMLS
- 4 techniques of increasing aggressiveness
  - Use Synonymy
  - Use Associated expressions (ATXs)
  - Explore the Ancestors
  - Explore the Other related concepts

[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? (MRCONSO)
If not,

Is there an associated expression (ATX) that describes this concept using a combination of MeSH descriptors? \((\text{MRATX/\text{MRMAP} + \text{MRREL}})\)

Endoscopic removal of intraluminal foreign body from oesophagus without incision

\(\text{MH/SH}\)

\(\text{Esophagus}\)

\(\text{surgery}\)

\(\text{Foreign Bodies}\)
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 (MRCONSO)

- 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 (MRCONSO)
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  Quantitative results

- 86% of UMLS concepts mapped to MeSH (2007)

**Other related concepts**
- Synonymy: 23%
- Graph of ancestors: 54%
- Built-in mappings: 11%
- Other: 12%
Restrict to MeSH Qualitative results

◆ Qualitative evaluation
  ● 1,036 concepts extracted from 200 MEDLINE citations
  ● manual review of every mapping or failure

◆ 61% Relevant
  ● Subtotal Gastrectomy ➔ Gastrectomy
  ● Encephalopathy, NOS ➔ Brain Diseases

◆ 28% More or less relevant
  ● Vitamin A measurement ➔ Laboratory Procedure
  ● Swelling, NOS ➔ Symptoms

◆ 11% Non relevant
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
Bibliography
References: UMLS home page

- UMLS home page

- UMLS documentation
  - Formerly known as the “Green Book”
  - Now online documentation
References

◆ Short presentation

◆ UMLS as a research project
References

◆ Technical papers

◆ Comprehensive bibliography 1986-96
  
Documentation and Support
UMLS documentation and support

- **UMLS homepage**
  - links to various UMLS resources

- **UMLSKS homepage**
  - links to the User’s and Developer’s guides

- **UMLS mailing list**
  - UMLSUSERS-L@LIST.NIH.GOV

- **Email address for support**
  - custserv@nlm.nih.gov