What does UMLS stand for?

- Unified
- Medical
- Language
- System

UMLS®
Unified Medical Language System®
UMLS Metathesaurus®
Motivation

- Started in 1986
- National Library of Medicine
- “Long-term R&D project”

«[...] 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.»
Overview through an example
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**

<table>
<thead>
<tr>
<th>eponym</th>
<th>symptoms</th>
<th>clinical</th>
<th>variants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Organize terms

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

<table>
<thead>
<tr>
<th>Term</th>
<th>MeSH</th>
<th>MeSH ID</th>
<th>MedDRA</th>
<th>MedDRA ID</th>
<th>ICD-10</th>
<th>ICD-10 Code</th>
<th>SNOMED CT</th>
<th>SNOMED CT ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addison Disease</td>
<td>MeSH</td>
<td>D000224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary hypoadrenalism</td>
<td>MedDRA</td>
<td>10036696</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary adrenocortical insufficiency</td>
<td>ICD-10</td>
<td>E27.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addison's disease (disorder)</td>
<td>SNOMED CT</td>
<td>363732003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0001403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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
Primary adrenocortical insufficiency

Disorders of other endocrine gland

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)
Relate to other concepts

- Additional hierarchical relations
  - link to other trees
  - make relationships explicit
- Non-hierarchical relations
- Co-occurring concepts
- Mapping relations
Endocrine Glands

Adrenal Glands

Adrenal Cortex

Adrenal Cortex Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Secondary hypocortisolism

Addison’s Disease

Addison’s disease due to autoimmunity

Diseases

Endocrine Diseases

Adrenal Dysfunction

Disorders of other endocrine gland

Other disorders of adrenal gland

relate to other concepts
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
  - Diseases
    - Endocrine Diseases
      - Adrenal Gland Diseases
        - Adrenal Gland Hypofunction
          - Addison’s Disease
UMLS Knowledge Sources
UMLS 3 components

- **Metathesaurus**
  - Concepts
  - Inter-concept relationships

- **Semantic Network**
  - Semantic types
  - Semantic network relationships

- **Lexical resources**
  - SPECIALIST Lexicon
  - Lexical tools
UMLS Metathesaurus
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

- 133 source vocabularies contributing concept names
- ~80 families of vocabularies
  - multiple translations (e.g., MeSH, ICPC, ICD-10)
  - variants (American-English equivalents, Australian extension/adaptation)
  - subsequent editions usually considered distinct families (ICD: 9-10; DSM: IIIR-IV)
- Broad coverage of biomedicine
- Common presentation (file format + Unicode)
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)
Addison’s Disease: Concept

A disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. It is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol. In the absence of replacement therapy, it is usually fatal.
Metathesaurus Concepts (2005AB)

- **Concept** (~ 1.2 M) CUI
  - Set of synonymous concept names
- **Term** (~ 4.2 M) LUI
  - Set of normalized names
- **String** (~ 4.8 M) SUI
  - Distinct concept name
- **Atom** (~ 5.6 M) AUI
  - Concept name in a given source

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0000001</td>
<td>headache</td>
<td>(source 1)</td>
</tr>
<tr>
<td>A0000002</td>
<td>headache</td>
<td>(source 2)</td>
</tr>
<tr>
<td>S0000001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0000003</td>
<td>Headache</td>
<td>(source 1)</td>
</tr>
<tr>
<td>A0000004</td>
<td>Headache</td>
<td>(source 2)</td>
</tr>
<tr>
<td>S0000002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0000001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0000005</td>
<td>Cephalgia</td>
<td>(source 1)</td>
</tr>
<tr>
<td>S0000003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0000002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0000001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cluster of synonymous terms

<table>
<thead>
<tr>
<th>Concept</th>
<th>S0001403</th>
<th>Term</th>
<th>L0001403</th>
<th>S0354372 Addison’s disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>S0010792 Addison Disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S0010794 Addison’s Disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S0010796 Addisons Disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S0033587 Disease, Addison</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S0352253 ADDISON’S DISEASE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Term</td>
<td>L2799243</td>
<td>S3341310 Addison’s disease (disorder)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S5907336 Primary Adrenocortical Insufficiency</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S5901878 Insufficiencies, Primary Adrenocortical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Term</td>
<td>L0494940</td>
<td>S5907334 Primary Adrenal Insufficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S4094828 adrenal; insufficiency, primary</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Term</td>
<td>L0494851</td>
<td>S5907343 Primary Hypoadrenalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>S5901432 Hypoadrenalism, Primary</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Term</td>
<td>L0585243</td>
<td>S1471573 Addison-Krankheit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GER</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Term</td>
<td>L1229627</td>
<td>S6107160 Maladie d'Addison</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FRE</td>
<td></td>
</tr>
</tbody>
</table>
Metathesaurus Evolution over time

- Concepts never die (in principle)
  - CUIs are permanent identifiers
- What happens when they do die (in reality)?
  - Concepts can merge or split
  - Resulting in new concepts and deletions

Addison's disease, NOS
C0271735

Addison's disease
C0001403


NLM
Metathesaurus Relationships

- Symbolic relations: ~9 M pairs of concepts
- Statistical relations: ~7 M pairs of concepts (co-occurring concepts)
- Mapping relations: 100,000 pairs of concepts

Categorization: Relationships between concepts and semantic types from the Semantic Network
Symbolic relations

◆ Relation
  ● Pair of “atom” identifiers
  ● Type
  ● Attribute (if any)
  ● List of sources (for type and attribute)

◆ Semantics of the relationship:
  defined by its type [and attribute]

Source transparency: the information is recorded at the “atom” level
Symbolic relationships

- **Hierarchical**
  - Parent / Child: PAR/CHD
  - Broader / Narrower than: RB/RN

- **Derived from hierarchies**
  - Siblings (children of parents): SIB

- **Associative**
  - Other: RO

- **Various flavors of near-synonymy**
  - Similar: RL
  - Source asserted synonymy: SY
  - Possible synonymy: RQ
Symbolic relationships

◆ Hierarchical
  ● isa (is-a-kind-of)
  ● part-of

◆ Associative
  ● location-of
  ● caused-by
  ● treats
  ● ...

◆ Cross-references (mapping)
UMLS Semantic Network
Semantic Network

- Semantic types (135)
  - tree structure
  - 2 major hierarchies
    - Entity
      - Physical Object
      - Conceptual Entity
    - Event
      - Activity
      - Phenomenon or Process
Semantic Network

- **Semantic network relationships (54)**
  - hierarchical (isa = is a kind of)
    - among types
      - Animal isa Organism
      - Enzyme isa Biologically Active Substance
    - among relations
      - treats isa affects
  - non-hierarchical
    - Sign or Symptom diagnoses Pathologic Function
    - Pharmacologic Substance treats Pathologic Function
"Biologic Function" hierarchy (isa)

Biologic Function

Physiologic Function
- Organism Function
  - Mental Process

- Organ or Tissue Function
  - Cell Function
    - Genetic Function

- Cell or Molecular Function
  - Mental or Behavioral Dysfunction

Pathologic Function
- Cell or Molecular Dysfunction
- Disease or Syndrome
  - Mental or Behavioral Dysfunction
- Experimental Model of Disease
  - Neoplastic Process
Associative (non-isa) relationships
Why a semantic network?

- Semantic Types serve as high level categories assigned to Metathesaurus concepts, *independently of their position in a hierarchy*

- A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs
  - The relationship may or may not hold at the concept level
  - Other relationships may apply at the concept level
Relationships can inherit semantics

![Semantic Network Diagram](image)

- Fully Formed Anatomical Structure
- Biologic Function
- Pathologic Function
- Body Part, Organ, or Organ Component
- Disease or Syndrome
- Adrenal Cortex
- Adrenal Cortical hypofunction
- Metathesaurus
SPECIALIST Lexicon and lexical tools
Lexical tools

To manage lexical variation in biomedical terminologies

Major tools

- Normalization
- Indexes
- Lexical Variant Generation program (lvgl)

Based on the SPECIALIST Lexicon

Used by noun phrase extractors, search engines
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

[Diagram showing normalization process]
Conclusions
Integrating subdomains

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

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

◆ UMLS
  umlsinfo.nlm.nih.gov

◆ UMLS browsers
  (free, but UMLS license required)
  ● Knowledge Source Server: umlsks.nlm.nih.gov
  ● Semantic Navigator:
  ● RRF browser
    (standalone application distributed with the UMLS)
Recent overviews


References

◆ UMLS as a research project


References

◆ Technical papers
