Unified Medical Language System® (UMLS®)

NLM Presentation Theater
MLA 2005
May 16 & 17, 2005

Rachel Kleinsorge
What does UML stand for?

- Unified
- Medical
- Language
- System

UMLS®
Unified Medical Language System®
The UMLS is:

Metathesaurus
- Concepts

Semantic Network
- Categories and relationships

SPECIALIST Lexicon & Lexical Tools
- Lexicon, lexical databases and programs

The Knowledge Sources
(delivered as machine readable files)
UMLS 3 Knowledge Sources

- **Metathesaurus**
  - Over 100 source vocabularies
  - Over 1 million concepts
  - Inter-concept relationships

- **Semantic Network**
  - 135 Semantic types (broad categories)
  - 54 Semantic relations (between categories)

- **Lexical resources**
  - SPECIALIST Lexicon
  - Lexical tools (programs and databases)
History of the UMLS

- Started at National Library of Medicine, 1986
- “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.]

[Lindberg & al., Methods, 1993]
[Humphreys & al., JAMIA, 1998]
UMLS Objectives

Knowledge Sources used to overcome:

- disparities in language format
  - Ex: atrial fibrillation, auricular fibrillation, af

- disparities in granularity and perspective
  - Ex: Contusions, hematoma, bruise

- problems in mapping and aggregating within and across databases and systems
  - Ex: Instruct patient to promptly report nosebleeds and excessive bruising (NIC), Epistaxis (MeSH)
UMLS in Practice

- Intellectual “middleware” = a set of multi-purpose tools for system developers
- Databases: 3 separate sets of relational files
- Tools:
  - MetamorphoSys (installation and customization)
  - Web interface Knowledge Source Server (UMLSKS)
  - Application programming interfaces
  - lvgl (lexical programs)
  - RRF Subset Browser

The UMLS is *not* an end-user application
UMLS Uses

- Information retrieval
- Thesaurus construction
- Natural language processing
- Automated indexing
- Electronic patient records
License Agreements

- Semantic Network, SPECIALIST Lexicon, and Lexical Programs
  - terms and conditions of use online

- Metathesaurus
  - license agreement process
  - some restrictions
    - 2. No charges, usage fees or royalties will be paid to NLM.
    - 5. Within 30 days of the end of any calendar year … provide NLM with a brief report
    - 11.c. required to include … identifiers from … the original source vocabularies
    - 12. For material … from some sources additional restrictions … may apply.
What is the UMLS?

*Overview through an example*
Addison’s Disease 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
  - eponym
  - symptoms
  - clinical
  - variants
Metathesaurus gathers and organizes terms

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

<table>
<thead>
<tr>
<th>Term</th>
<th>System</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenal gland diseases</td>
<td>MeSH</td>
<td>D000307</td>
</tr>
<tr>
<td>Adrenal disorder</td>
<td>AOD</td>
<td>00000005418</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>

C0001621

Adrenal Gland Diseases
## Cluster of synonymous terms

<table>
<thead>
<tr>
<th>Concept</th>
<th>C0001621</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>L0001621</td>
</tr>
<tr>
<td>Term</td>
<td>L0041793</td>
</tr>
<tr>
<td>Term</td>
<td>L0161347</td>
</tr>
<tr>
<td>Term</td>
<td>L0181041</td>
</tr>
<tr>
<td>Term</td>
<td>L0368399</td>
</tr>
<tr>
<td>Term</td>
<td>L1279026</td>
</tr>
<tr>
<td>Term</td>
<td>L0162317</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term ID</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0011232</td>
<td>Adrenal Gland Diseases</td>
</tr>
<tr>
<td>S0011231</td>
<td>Adrenal Gland Disease</td>
</tr>
<tr>
<td>S0000441</td>
<td>Disease of adrenal gland</td>
</tr>
<tr>
<td>S0481705</td>
<td>Disease of adrenal gland, NOS</td>
</tr>
<tr>
<td>S0220090</td>
<td>Disease, adrenal gland</td>
</tr>
<tr>
<td>S0044801</td>
<td>Gland Disease, Adrenal</td>
</tr>
<tr>
<td>S0860744</td>
<td>Disorder of adrenal gland, unspecified</td>
</tr>
<tr>
<td>S0217833</td>
<td>Unspecified disorder of adrenal glands</td>
</tr>
<tr>
<td>S0225481</td>
<td>ADRENAL DISORDER</td>
</tr>
<tr>
<td>S0627685</td>
<td>DISORDER ADRENAL (NOS)</td>
</tr>
<tr>
<td>S0632950</td>
<td>Disorder of adrenal gland</td>
</tr>
<tr>
<td>S0354509</td>
<td>Adrenal Gland Disorders</td>
</tr>
<tr>
<td>S0586222</td>
<td>Adrenal disease</td>
</tr>
<tr>
<td>S0466921</td>
<td>ADRENAL DISEASE, NOS</td>
</tr>
<tr>
<td>S1520972</td>
<td>Nebennierenkrankheiten</td>
</tr>
<tr>
<td>S0226798</td>
<td>SURRENALE, MALADIES</td>
</tr>
</tbody>
</table>

[...]

[...]

[...]
Semantic Network

- Organizes terms into 135 broad subject categories
  - Semantic Types (Clinical Drug, Virus)
  - Addison’s Disease **Semantic Type: Disease or Syndrome**

- Defines 54 Semantic Relationships
  - Links between categories (isa, causes, treats)
  - Ex: Virus **causes** Disease or Syndrome

- Together, types and relations:
  - Form the structure of the semantic network
  - Broadly categorize the biomedical domain
Lexical Tools
process terms and text

Organization is accomplished using:
- Normalization
- Semantic pre-processing
- UMLS editors

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

Metathesaurus:
◆ clusters terms into concepts – assigns unique identifier

Semantic Network:
◆ defines relationships between concepts, organizes concepts into categories

Lexicon and Lexical Tools:
◆ process terms for entry into the Metathesaurus