The Unified Medical Language System
What is it and how to use it?

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

Outline

◆ What is the UMLS?
  ● Introduction
  ● Overview through an example
  ● The three UMLS Knowledge Sources

◆ How to use the UMLS?
  ● Obtaining a license
  ● Remote access
  ● Local installation and customization
  ● A UMLS-based algorithm
  ● Benefits and limitations

Part I
What is the UMLS?

(1) Introduction

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”
- 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 need for the same concept to be expressed in different machine-readable sources and by different people.
- The second is the need for 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
  - Is (lexical programs)
  - MetamorphoSys (installation and customization)

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

Part I
What is the UMLS?

(2) Overview through an example

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 adrenocortical insufficiency
  - Chronic adrenocortical insufficiency
- Contexts: different hierarchies

Organize terms

- 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>Adrenal Gland Diseases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SNOMED International

Diseases/Diagnoses

Diseases of the endocrine system

Diseases of the Adrenal Glands

Addison’s Disease

MeSH

Diseases

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s Disease

AOD

Endocrine disorder

Adrenal disorder

Adrenal cortical disorder

Adrenal cortical hypofunction

Addison’s Disease

Read Codes

Endocrine disorder

Disorder of adrenal gland

Hypoadrenalism

Adrenal Hypofunction

Corticoadrenal insufficiency

Addison’s Disease
Primary adrenocortical insufficiency

Other disorders of adrenal gland

Disorders of other endocrine gland

Inter-concept relationships: hierarchies from the source vocabularies

Redundancy: multiple paths

One graph instead of multiple trees (multiple inheritance)

Organize concepts

Adrenal Cortex Diseases

Addison’s Disease

Adrenal Gland Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Endocrine Diseases

Adrenal Gland Dysfunction

Non-hierarchical relationships

Co-occurring concepts

Mapping relationships

Endocrine System

Abdominal organ

Adrenal Glands

Adrenal Cortex Diseases

Adrenal Cortex Hypofunction

Adrenal Gland Hypofunction

Secondary hypofunction

Adrenal’s Disease

Categorize concepts

Endocrine Diseases

Addison’s Disease

Disease or Syndrome

High-level categories (semantic types)

Assigned by the Metathesaurus editors

Independently of the hierarchies in which these concepts are located

Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s disease due to autoimmune

Relate to other concepts
How do they do that?

- Lexical knowledge
- Semantic pre-processing
- UMLS editors

Lexical knowledge

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

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 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 (2005AA)

- 134 source vocabularies
- 132 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: III-R-IV)
- Broad coverage of biomedicine
- 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)
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

- Concept (~ 1.2M) CUI
  - Set of synonymous concept names
- Term (~ 4.2 M) LUI
  - Set of normalized names
- String (~ 4.7M) SUI
  - Distinct concept name
- Atom (~ 5.5M) AUI
  - Concept name in a given source

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

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
  - Broader / Narrower than

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

- **Associative**
  - Other

- **Various flavors of near-synonymy**
  - Similar
  - Source asserted synonymy
  - Possible synonymy

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 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
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

SPECIALIST Lexicon

- Content
  - English lexicon
  - Many words from the biomedical domain
- 200,000+ lexical items
- Word properties
  - morphology
  - orthography
  - syntax
- Used by the lexical tools
Morphology

- **Inflection**
  - noun: nucleus, nuclei
  - verb: cauterize, cauterizes, cauterized, cauterizing
  - adjective: red, redder, reddest

- **Derivation**
  - verb ← noun: cauterize → cauterization
  - adjective ← noun: red → redness

Orthography

- **Spelling variants**
  - oe/e: oesophagus - esophagus
  - ae/e: anaemia - anemia
  - is/e/z: cauterise - cauterize
  - genitive mark: Addison's disease

Syntax

- **Complementation**
  - verbs
    - intransitive: I'll treat.
    - transitive: He treated the patient.
    - ditransitive: He treated the patient with a drug.
  - nouns
    - prepositional phrase: Valve of coronary sinus

Lexical tools

- **To manage lexical variation in biomedical terminologies**
- **Major tools**
  - Normalization
  - Indexes
  - Lexical Variant Generation program (lvg)
- **Based on the SPECIALIST Lexicon**
- **Used by noun phrase extractors, search engines**

Normalization

- **Example**
  - Hodgkin's disease, NOS
  - Hodgkin diseases, NOS
  - Hodgkin disease, NOS
  - Disease, Hodgkin
  - Hodgkin disease, NOS
  - Hodgkin's disease
  - Hodgkin Diseases
  - Disease, Hodgkin
  - Disease, Hodgkin
  - Hodgkin's disease
  - Hodgkin Disease
  - Hodgkin's Disease
  - Hodgkin's Disease
  - Disease, Hodgkin
  - Disease, Hodgkin
  - Disease, Hodgkin
  - Disease, Hodgkin

Normalization: Example

- Hodgkin's disease, NOS
- Hodgkin diseases, NOS
- Hodgkin disease, NOS
- Disease, Hodgkin
- Hodgkin disease, NOS
- Hodgkin's disease
- Hodgkin Diseases
- Disease, Hodgkin
- Disease, Hodgkin
- Disease, Hodgkin
- Disease, Hodgkin
- Hodgkin's disease
- Hodgkin Disease
- Hodgkin's Disease
- Hodgkin's Disease
- Disease, Hodgkin
- Disease, Hodgkin
- Disease, Hodgkin
- Disease, Hodgkin
- Hodgkin's disease
- Hodgkin Disease
- Hodgkin's Disease
- Hodgkin's Disease
- Disease, Hodgkin
- Disease, Hodgkin
- Disease, Hodgkin
- Disease, Hodgkin
Normalization Applications

- Model for lexical resemblance
- Help find lexical variants for a term
  - Terms that normalize the same usually share the same LUI
- Help find candidates to synonymy among terms
- Help map input terms to UMLS concepts

Indexes

- Word index
  - word to Metathesaurus strings
  - one word index per language
- Normalized word index
  - normalized word to Metathesaurus strings
  - English only
- Normalized string index
  - normalized term to Metathesaurus strings
  - English only

Lexical Variant Generation program

- Tool for specialists (linguists)
- Performs atomic lexical transformations
  - generating inflectional variants
  - lowercase
  - …
- Performs sequences of atomic transformations
  - a specialized sequence of transformations provides the normalized form of a term (the norm program)

Part II

How to use the UMLS?

Outline

- Part II: How to use the UMLS?
  - Obtaining a license
  - Remote access
    - Knowledge Source Server (UMLSKS)
    - UMLSKS Application programming interface (API)
  - Local installation and customization (MetamorphoSys)
  - A UMLS-based algorithm: Restrict to MeSH
  - Benefits and limitations

Part II

How to use the UMLS?

(1) Obtaining a license
First step: License agreement

- Online Web-based license:
  
  [link]

- Read license
- Read appendix
- Print a copy for your records
- Complete the Web form

- Verify:
  - receive e-mail from NLM to Web site within 72 hours and enter first and last name
  - NLM official will countersign (turn-around time of a few days)

- Receive 2nd e-mail from NLM with new license number

---

[link]

---

---
License Restriction Levels 0-4 (2004AB)

- Level 0 (28.2%)
  - unrestricted

- Level 1 (1.6%)
  - negotiate to translate

- Level 2 (0.4%)
  - negotiate to use in health data creation

- Level 3 (30.6%)
  - negotiate to use in production
  - explicitly prohibited to provide Internet access

- Level 4 (39.2%)
  - unrestricted for U.S. use and distribution

Remote Access

- Web search interface
- Application Programming Interface (API)
Knowledge Source Server

Web search interface

UMLSKS Web search interface

- Logging in
- Basic searching
- Advanced searching

UMLSKS Web search interface log in

Returning users log in
New users create account

UMLS Knowledge Source Server Home Page

- Tabs across top access basic searching of 3 Knowledge Sources
- Advanced searching options on right-hand side

UMLS Knowledge Source Server Home Page

Metathesaurus Basic Search
Addison’s disease

- UMLS Release
- Search Term
- UMLS Knowledge Source
Addison’s disease in SNOMED CT
Preferred Term and Code

<table>
<thead>
<tr>
<th>UMLS Knowledge Source Server (UMLSKS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metathesaurus Advanced Search Feature</td>
</tr>
<tr>
<td>Relational Record Request</td>
</tr>
</tbody>
</table>

- **TTY:** Term Type
- **ID:** Source Code Descriptor

Relational Records MRCONSO.RRF

Semantic Network Searching

- Select Tab along top
- Quick search
- Advanced Search on right-hand side

Semantic Network Search

- Enter search string
- Select semantic type
- Select semantic relation

Semantic Type Clinical Drug

- Browse ST hierarchy
- View Concepts with ST
- View Relations valid for the ST
- View Raw Relational Records
Show Relations Between Types

Validates whether a selected Semantic Relationship (SR) holds between two selected Semantic Types (ST)

SPECIALIST Lexicon Searching

Select Tab along top

Quick search

SPECIALIST Lexicon Search

SPECIALIST Lexical Record

UMLS Resources

NLP & Lexical Resources
- MetaMap Transfer (MMTs)
- Word Sense Disambiguation (WSD) Test Collection

Semantic Network
- Semantic Navigator
- Semantic Groups
- Metathesaurus
- String Properties
Knowledge Source Server
Application Programming Interface

UMLSKS API basics
- Remote server at NLM
- Local application connected through
  - Java RMI
    - Java-based applications
    - Developer’s Guide: Chapter 3
    - Set of Java classes (part of the UMLSKS API)
    - Detailed Javadoc documentation online and with API download
  - TCP/IP socket
    - XML-based queries
    - Developer’s Guide: Chapter 5
    - XML schema
    - Socket server
      - Host: umlsks.nlm.nih.gov
      - Port: 8042

Developer’s Guide
- User’s Guide
- Developer’s Guide
  - Java RMI
  - XML-based queries
  - Socket server
    - Host: umlsks.nlm.nih.gov
    - Port: 8042

Documentation Javadocs

Sample XML query (1) Current version
```xml
<?xml version="1.0"?>
<getCurrentUMLSVersion version="1.0"/>
<?xml version="1.0"?>
<CurrentUMLSVersion version="1.0">
  2004AB
</CurrentUMLSVersion>
```
Sample XML query (2) Concepts by string

```xml
<?xml version="1.0" ?>
<findCUI version="1.0">
  <conceptName>appendicectomy</conceptName>
  <language>ENG</language>
  <exact/>
  <noSuppressibles/>
</findCUI>
```

Sample XML query (3) Concepts properties

```xml
<?xml version="1.0" ?>
<ConceptIdCollection version="1.0">
  <release>2004AB</release>
  <conceptId>
    <cui>C0003611</cui>
    <cn>Appendectomy</cn>
  </conceptId>
</ConceptIdCollection>
```

Sample XML query (4) Relationships

```xml
<?xml version="1.0" ?>
<getRelations version="1.0">
  <cui>C0033572</cui>
  <rel>RO</rel>
</getRelations>
```

Sample XML query (5) All semantic type IDs

```xml
<?xml version="1.0" ?>
<listSemTypeIds version="1.0">
</listSemTypeIds>
```

Performing XML queries from UMLSKS

```xml
<?xml version="1.0" ?>
<getRelations>
  <cui>C0033572</cui>
  <rel>RO</rel>
</getRelations>
```
Part II
How to use the UMLS?

(3) Installing the UMLS locally and Customizing the Metathesaurus using MetamorphoSys

What is MetamorphoSys?
- Tool distributed with the UMLS
- Multi-platform Java software
- The UMLS installation and customization wizard
  - Installs Knowledge Sources to local storage
  - Subsets and customizes a local Metathesaurus

Using MetamorphoSys
- Simple to use
- Screens and tabs lead you through process
- Installs NLM data format files to local storage

Why use MetamorphoSys?
Customize the Metathesaurus
- To remove terminology that is unhelpful, or even harmful, to your needs and purposes
- To comply with terms of license agreement

Why use MetamorphoSys?
Changing Default Settings
- To alter the preferred name
- To alter suppressibility of specific source term types

Customization is Critical
- Requires a clear understanding of:
  - Characteristics of source vocabularies
  - License arrangements
  - User’s functional requirements
  - User’s purpose and perspective
- Technical expertise
  - ... and requires a multidisciplinary technical team
Machine Requirements

- A fast CPU – 1 GHz or higher
- 1 GB RAM recommended (512 MB min.)
- 6x (or better) DVD drive
- 22 GB minimum free disk space
- Runs on Sun Solaris 8 & 9, Windows XP, NT, and 2000, Linux, and Mac
- 1-10 hours run time on platforms tested

Download from UMLSKS …

- High speed Internet connection required
- Read the README file for the release
- 2004AB UMLS Files
  - 2004ab.CHK
  - 2004ab.MD5
  - 2004ab-1-meta.nlm
  - 2004ab-2-meta.nlm
  - 2004ab-3-meta.nlm
  - mmsys.zip
  - Copyright_Notice.txt
  - README.txt
  - Please README!

...or DVD?

- Order at: umls_support@nlm.nih.gov
- Include your license number
- Run MetamorphoSys from DVD
  - Windows: Autorun; or go to root directory and click on “windows_mmsys.bat”
  - Linux, Solaris, Macintosh: open a terminal window, change to the root directory and type appropriate command: ./linux_mmsys.sh, ./solaris_mmsys.sh, ./macintosh_mmsys.sh

Be patient! A lot of software must load.

Welcome Screen

Install UMLS
Install UMLS Advanced Options

- Copy NLM data format files (.zrc) to hard drive
- Copy MetamorphSys to hard drive
- Run MG5 validation

Done  Cancel

UMLS License Notice

[Image of license agreement]

Installation progress monitor

Select a default subset

Level 0 → no separate additional license agreements
Level 0 + SNOMED CT → Non-U.S. users must have separate license agreements
RxNorm → no separate additional license agreements

Input Options Tab

- Customize the input of UMLS data. See Help for more information.

Output Options Tab

- Select data output options for your local application. See Help for more information.
Source List Tab

MetamorphoSys Option Tab

Precedence Tab

Suppressibility Tab

File menu
Edit menu

Options menu

Reset menu

- Returns all filters to default selections
- Default selections in “mmsys.prop.default file” in config folder
- mmsys.prop.default contains properties in last run
How MetamorphoSys Works

- Removes all information from relational files in excluded vocabularies
  - atoms, strings, relationships, attributes, mappings, etc.
- Applies additional options selected by user
  - such as adding source term suppressibility or altering precedence
- Produces a full set of Metathesaurus files
  - relational files with customized data
  - reflecting other user criteria

MetamorphoSys log

Kept Sources

Output directory contents
Part II
How to use the UMLS?

(4) A UMLS-based algorithm

Indexing Initiative
[Aronson et al., AMIA, 2000]

- 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

UMLS

Restrict to MeSH
[Bodenreider et al., AMIA, 1998]

- 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

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)

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

Restrict to MeSH  Example

Vein of neck, NOS
Vein of head and neck, NOS

Neck
Blood Vessels
Vascular structure

Restrict to MeSH  Quantitative results

- 82.5% of UMLS concepts mapped to MeSH

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

Part II
How to use the UMLS?

(5) Benefits and Limitations
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

UMLA as an enabling resource

- Examples
  - Mapping across vocabularies
  - Semantics of statistical associations
  - Redundancy in hierarchical relations

Limitations

- Structural inconsistency
  - Cycles in the graph of hierarchical relations
- Semantic inconsistency
  - Between Metathesaurus and Semantic Network
- Missing relations
  - Synonymy
  - Hierarchical relations (missing or underspecified)

[Cimino, JAMIA, 1998]
Structural inconsistency  From trees to graph

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

Structural inconsistency  There are some cycles

Semantic inconsistency  A two-level structure

Semantic inconsistency  A limited study

Semantic inconsistency  Issues

- The UMLS integrates what terminologies represent
- Hierarchies in source vocabularies
  - Often task-driven rather than based on principles
  - Usually suitable for information retrieval
  - Not necessarily suitable for reasoning
- No automatic correction possible
  - Wrong categorization
  - Wrong inter-concept relationship
  - [Wrong semantic network relationship]
Missing relations  Example

- diseases of the skin and subcutaneous tissues
- acute eczema
- infantile eczema
- eczema
- acute infantile eczema

Missing relations  Example

- diseases of the skin and subcutaneous tissues
- acute eczema
- infantile eczema
- eczema
- acute infantile eczema

Missing relations  A limited study

- 28,851 pairs of terms
  - Original SNOMED term
  - Demodified 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

[Bohdenreider & al., TIA, 2001]

Compensation mechanisms

- Examples
  - Removing cycles from hierarchical relations
  - Using redundancy (number of sources asserting the relation)
  - Using terminological knowledge (e.g., NEC)
  - Lexically-suggested hyponymic relations
  - Properties of adjectival modification

More limitations

- Meaning of isa
- Some missing / wrong relations are hard to detect
- Some relations are present but hard to find

Meaning of isa

- Autoimmune Diseases
- Addison’s disease
- Tuberculosis
- Addison’s disease (due to autoimmunity)
Relations Missing and difficult to detect

- chronic hypertensive renal failure
- chronic renal failure
- hypertensive renal failure

Relations Existing but difficult to find

- ferritin
- iron ion
- iron ion
- transport
- has function
- ferritin
- carrier protein
- iron
- iron-binding protein

UMLS Gene Ontology
- ferritin isa
- iron transporter ferritin
- transports
- iron
- reified "transport" relationship
- "transport" relationship

How to address these limitations?

- Description logics
- Natural Language Processing (semantic interpretation of the terms)
- Comparing knowledge sources (alignment, inference)

Summary

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

Medical Ontology Research

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

UMLS Overview

- UMLS = 3 Knowledge Sources
  - Metathesaurus
  - Semantic Network
  - SPECIALIST Lexicon and Lexical Tools
- MetamorphoSys
  - installs
  - customizes
- UMLSKS
  - remote access
  - resources and documentation
Bibliography

References

- UMLS as a research project
- Short presentation

References

- Technical papers
- Comprehensive bibliography 1986-96

UMLS documentation and support

  - with links to all other UMLS information
  - with links to the User’s and Developer’s guides
- Email address for support custserv@nlm.nih.gov
## Appendix 1

### UMLS files in Rich Release Format

#### MRCONSO (sample rows 1..5)

<table>
<thead>
<tr>
<th>CUI</th>
<th>LAT</th>
<th>LUI</th>
<th>SUI</th>
<th>AUI</th>
<th>SAUI</th>
<th>SCUI</th>
<th>SDUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>D000224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### MRCONSO (sample rows 6..10)

<table>
<thead>
<tr>
<th>CUI</th>
<th>LAT</th>
<th>LUI</th>
<th>SUI</th>
<th>AUI</th>
<th>SAUI</th>
<th>SCUI</th>
<th>SDUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>D000224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### MRCONSO (sample rows 11-13)

<table>
<thead>
<tr>
<th>CUI</th>
<th>LAT</th>
<th>LUI</th>
<th>SUI</th>
<th>AUI</th>
<th>SAUI</th>
<th>SCUI</th>
<th>SDUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>D000224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### MRHIER (sample rows)

<table>
<thead>
<tr>
<th>CUI</th>
<th>AUI</th>
<th>CUN</th>
<th>PASS</th>
<th>SAB</th>
<th>RELA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### MRREL (sample rows)

<table>
<thead>
<tr>
<th>CUI</th>
<th>SYSTYPE</th>
<th>REL</th>
<th>CUI</th>
<th>SYSTYPE</th>
<th>REL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Section titles may vary depending on the specific context or application within the UMLS files.
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.
Addison's Disease

- Adrenal Gland Hypofunction
- Endocrine Diseases
- Diseases (MeSH Category)
- Hypoaldosteronism
- Index Medicus Descriptor
- MeSH Descriptors
- MeSH

Primary adrenocortical insufficiency

- Other disorders of adrenal gland
- Disorders of other endocrine glands
- ICD…, Tenth Revision (ICD-10)

Addison's disease, NOS

- Diseases of the Adrenal Glands
- Diseases of the End. System
- Diseases/Diagnoses
- SNOMED International

Addison's disease

- Other diseasess of the adrenal gland
- Disorders of other endocrine glands
- ICD…, Tenth Revision (ICD-10)
SRDEF  Basic information  (2003AA)

| SRDEF | SRDEF: Basic information | ...
|-------|--------------------------|-----
| Def    | Generally, a living individual, including all plants and animals. (Assign to Organism if organism only.) | Organism
| Def    | Plant: An autotroph (photosynthetic organism) with a cellular structure, and lacking the power of locomotion. | Plant
| Def    | Alga: A chiefly aquatic plant that contains chlorophyll, but does not form embryos during development and lacks vascular tissue. | Alga
| Def    | Fungus: An eukaryotic organism characterized by the absence of chlorophyll and the presence of a rigid cell wall. | Fungus

Appendix - Semantic Network relational files (ORF)  223

SRSTR  Structure  (2003AA)

| SRSTR | SRSTR: Structure | ...
|-------|-----------------|-----
| SRSTR | Disease or Syndrome | Disease or Syndrome
|        | Affects | Affects
|        | Organism | Organism

Appendix - Semantic Network relational files (ORF)  224

SRSTRE2  Structure (expanded)  (2003AA)

| SRSTRE2 | SRSTRE2: Structure (expanded) | ...
|---------|-----------------------------|-----
| SRSTRE2 | Disease or Syndrome | Disease or Syndrome
|         | Affects | Affects
|         | Organism | Organism

Appendix - Semantic Network relational files (ORF)  225