

Nucleic Acids Research 32 (267-270)

The Unified Medical Language System (UMLS):
integrating biomedical terminology

Olivier Bodenreider

Lister Hill Center for Biomedical Communications, National Library of Medicine, National
Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA

04/11/2004

Leo Wolf, ISiLS-UMLS

1

UMLS: Introduction

- Project started by National Library of Medicine (90's)
 - NLM: part of US Department of Health and Human Services
- Collaboration between:
 - Physicians, Computer Scientists, Linguists and research groups
- Purpose: intelligent information retrieval
"facilitate the development of computer systems that behave as if they "understand" the meaning of the language of biomedicine and health"
- Target audience:
 - developers: retrieve and integrate biomedical & health data
 - informatics researchers
- Result: Knowledge Sources and Tools (CD-ROM)

04/11/2004

Leo Wolf, ISiLS-UMLS

Knowledge
Source Server:
API + Web
interface

1. Knowledge Sources

Available resources:

- [Metathesaurus®](#)
- Semantic Network
- SPECIALIST Lexicon

Multi-purpose in nature; about concepts used in:

- patient records and public health data
- scientific literature
- guidelines

04/11/2004

Leo Wolf, ISiLS-UMLS

3

UMLS Metathesaurus, 1

- Multi-purpose, multi-lingual vocabulary database
- Concepts, names and relationships between them
- Input from ([source vocabularies](#)):
 - many different thesauri
 - classifications
 - code sets
 - lists of controlled termsused in:

patient care, health services billing, public health statistics,
indexing and cataloging biomedical literature and research

04/11/2004

Leo Wolf, ISiLS-UMLS

4

Used biomedical vocabularies (60)

- Systemetized Nomenclature of Medicine: SNOMED
 - International Classification of Diseases: ICD (10th since 1900)
 - Universal Medical Device Nomenclature System
 - WHO Adverse Drug Reaction Terminology
 - Physicians' Current Procedural Terminology: CPT
 - Medical Subject Headings: MeSH (index for MEDLINE)
 - Diagnostic and Statistical Manual of Mental Disorders: DSMIV
 - Thesaurus of Psychological Index Terms
- specifically for bio-scientists:
- NCBI taxonomy (identification of organisms)
 - Gene Ontology (gene product annotation)
 - Digital Anatomist Symbolic Knowledge Base

04/11/2004

Leo Wolf, ISiLS-UMLS

5

IDC-9-CM example

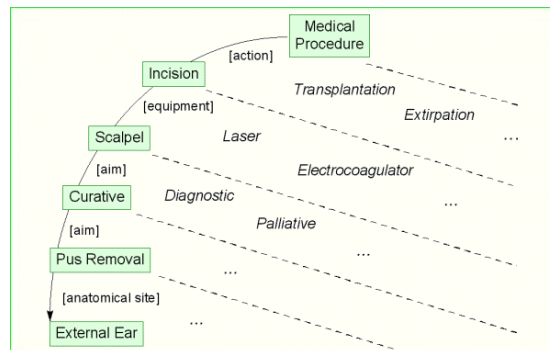
<i>Code</i>	<i>Disease</i>
001 - 139	Infectious and parasitic diseases
001 - 009	Infectious diseases of the digestive tract
003	Other Salmonella Infections
- 003.0	Salmonella gastroenteritis
- 003.1	Salmonella Septicemia
- 003.2	Localized Salmonella Infections
- 003.20	Localized Salmonella Infection, Unspecified
- 003.21	Salmonella Meningitis
- 003.22	Salmonella Pneumonia
- 003.23	Salmonella Arthritis
- 003.24	Salmonella Osteomyelitis
- 003.29	Other Localized Salmonella Infections
- 003.8	Other Specified Salmonella Infections
- 003.9	Salmonella Infections, Unspecified

04/11/2004

Leo Wolf, ISiLS-UMLS

6

Multi-axial classification

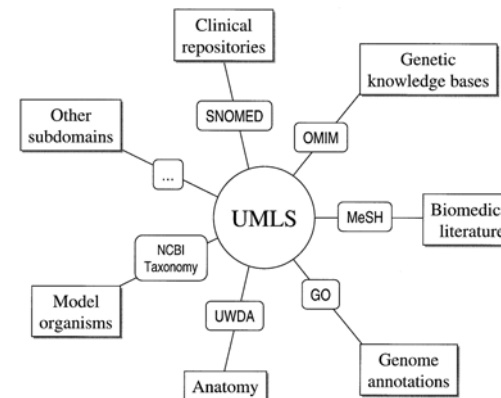


04/11/2004

Leo Wolf, ISiLS-UMLS

7

UMLS, Terminologies and subdomains



04/11/2004

Leo Wolf, ISiLS-UMLS

8

UMLS Metathesaurus, 2

- Organized by concepts (> 900k)
- Links alternative names (2.5M) of one concept:
 - synonyms, lexical variants and translations
- Identifies relationships (12M) between concepts
- Each concept is semantically typed at least once
- Concepts represented in Semantic Network
- Also linked to Lexicon
- Input source encoding is stored with concepts
- both files (Rich Release Format, RRF) and relational tables (Oracle DBMS)
- CUI <=> (LUI) => SUI => AUI (=> 1:M, <=> M:N)

04/11/2004

Leo Wolf, ISiLS-UMLS

9

UMLS Metathesaurus, 3

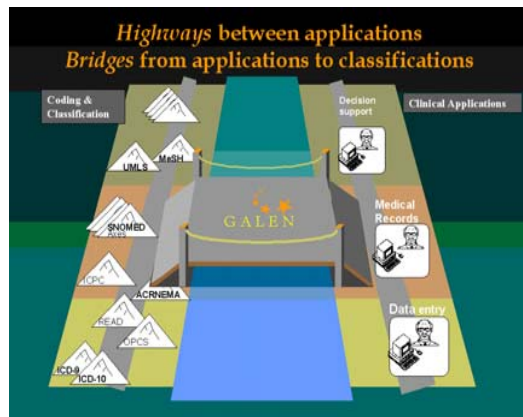
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation (preferred) Atrial Fibrillations	S0016668 Atrial Fibrillation (preferred)	A0027665 Atrial Fibrillation (from MSH) A0027667 Atrial Fibrillation (from PSY)
		S0016669 Atrial Fibrillations	A0027668 Atrial Fibrillations (from MSH)
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	S0016899 Auricular Fibrillation (preferred)	A0027930 Auricular Fibrillation (from PSY)
		S0016900 (plural variant) Auricular Fibrillations	A0027932 Auricular Fibrillations (from MSH)

04/11/2004

Leo Wolf, ISiLS-UMLS

10

Related project: GALEN



04/11/2004

Leo Wolf, ISiLS-UMLS

11

UMLS Semantic Network, 1

- Uses information from Metathesaurus
- Provides consistent categories (135) for all concepts and relationships (54)
- Defines semantic categories like:
 - organism
 - anatomical structure and biological function
 - chemicals and physical objects
- Relationships:
 - hierarchical: is_a, part_of, is_a_kind_of
 - associative: location_of, caused_by

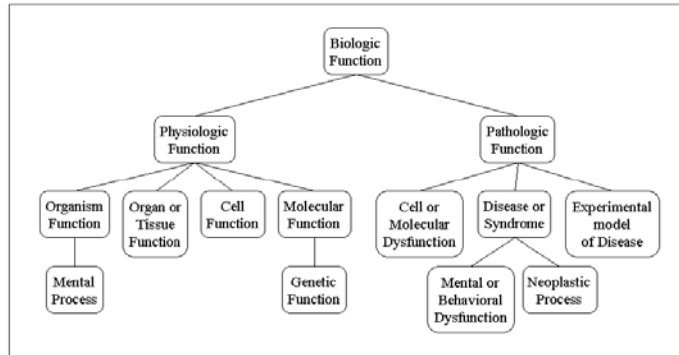
04/11/2004

Leo Wolf, ISiLS-UMLS

12

UMLS Semantic Network, 2

“Biologic function” hierarchy, lines depict is_a relationships

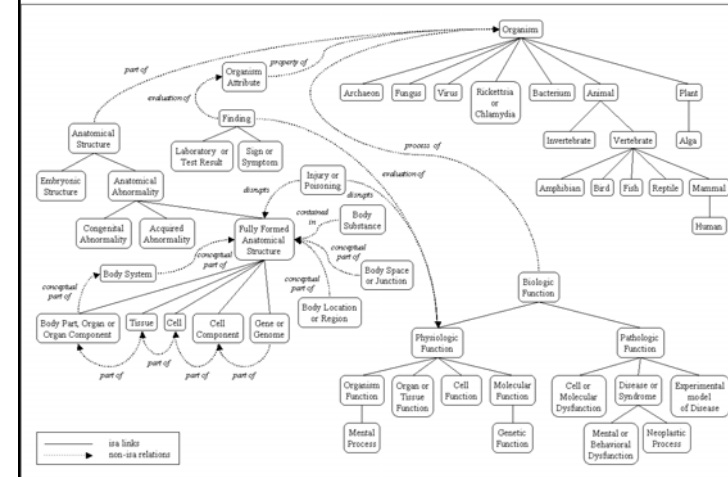


04/11/2004

Leo Wolf, ISIIS-UMLS

13

UMLS Semantic Network, 3



UMLS Example

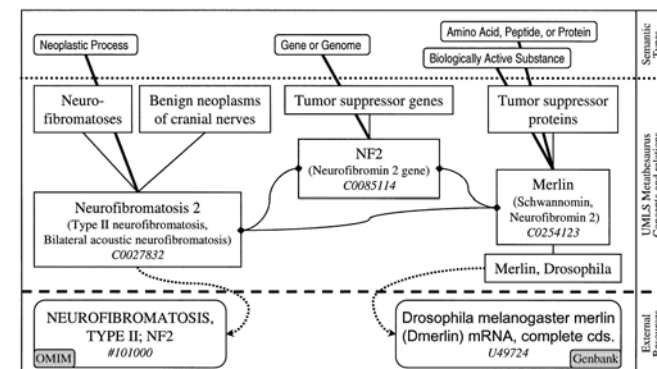
1. Neurofibromatosis 2, autosomal disease, causing tumors involving acoustic nerves. Caused by mutation in gene NF2, leading to diminished (or absence of) protein production
2. Found in Metathesaurus:
 - concepts: Neurofibromatosis 2, NF2, Merlin (proteine)
 - many synonyms
 - associative relationships
 - hierarchical relationships
3. Addition of UMLS semantic types leads to: [fig2](#)

04/11/2004

Leo Wolf, ISIIS-UMLS

15

NF2, related proteins and diseases (partial)



04/11/2004

Leo Wolf, ISIIS-UMLS

16

UMLS KSS Developers issues

- Set of machines, programs (Java) and API's ([Java](#))
- Accessed via:
 - Web: HTTP client (browser)
 - Java API (RMI)
 - non-Java: socket communication

Support for:

- XML input/output
- XML based queries: embedded metadata

Downloadable:

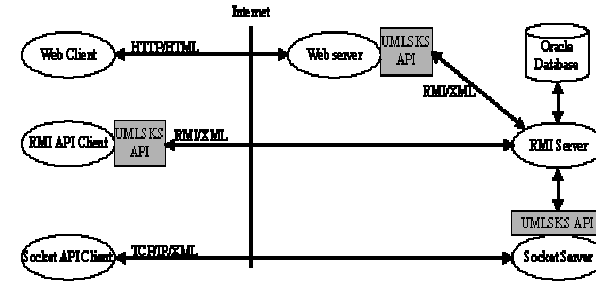
- relational files => SQL queries
- tools

04/11/2004

Leo Wolf, ISIILS-UMLS

17

KSS logical architecture

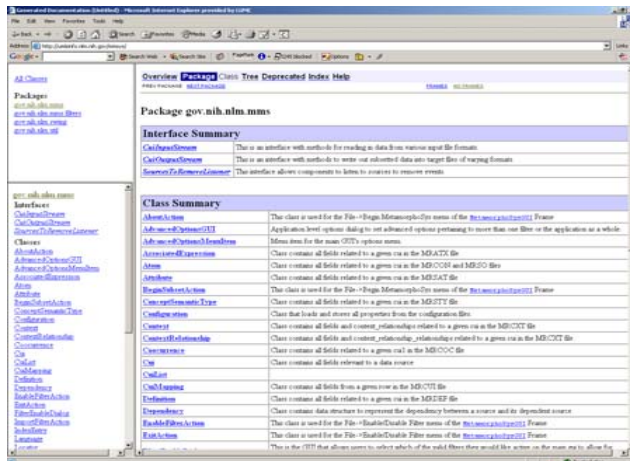


04/11/2004

Leo Wolf, ISIILS-UMLS

18

Java API, packages



2. Tools

1. MetamorphoSys

cross-platform Java program; for local installation of:

- Metathesaurus: customization (create/refine subsets)
- Semantic Network
- SPECIALIST Lexicon (NLP support for English)

2. Lexical Programs

"Ivg", to cope with variability in natural languages like:

- inflection: treat (verb) -> treats, treated, treating
- word order variants
- spelling and character set variants

04/11/2004

Leo Wolf, ISIILS-UMLS

20