



Terminological Resources in the Medical Domain

Pierre Zweigenbaum¹

¹LIMSI, CNRS, Université Paris-Saclay, Orsay, France

MiRoR, Split Workshop



- 1 Medical Terminologies
 - Medical Terminologies: Why They Are Needed
 - Terminologies: General Properties
- 2 Some Terminologies in Use in the Medical Domain
 - MeSH, ICD
 - SNOMED
 - WHO-ART, MedDRA, LOINC, Procedures
- 3 Terminology Unification: the UMLS Metathesaurus
 - Metathesaurus
 - Semantic Network
 - UMLS “Specialist” Tools
- 4 Conclusion

Representing Medical Information

Need to “compute” with medical information

- Counting, aggregation, statistics
 - epidemiology, medico-economical analysis
- Information retrieval in databases and knowledge-bases
 - access to scientific literature
- Accessing information for a given task
 - consulting an in-patient’s file during a hospital stay

→ Need to record medical information

Natural Language?

Natural language, or free text:

- contrasts with formal languages
- is ubiquitous, including in health
- is powerful and flexible: an asset for human expression
- but an obstacle to information processing

Need for a *normalized representation* of medical information

Information Exchange

Communication between two persons or computers:

sender



receiver

- Preserve integrity of message;
- Preserve meaning of message:

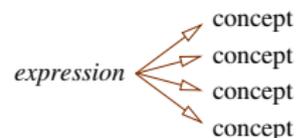
Give the same interpretation to the message

The issue

No perfect one-to-one (bijective) relationship between natural language expressions and domain concepts

Not Bijective: Ambiguity

Lack of consensus: variation between places, variation between periods



Polysemy: word with several senses

- *iris* (AOD, LNC, MeSH, CSP, NCI, RCD, ...) → a part of the eye
- *iris* (SNOMED CT, NCBI) → a flower
- *sinus* (CCPSS, CST, LNC, SNOMED CT) → nasal sinus
- *sinus* (SNOMED CT) → fistula
- *sinus* (MMSL) → acetaminophen-pseudoephedrine brand
- *sinus* (NCI, UWDA) → general anatomical term

Lack of precision: insufficiently specified description

- *infarction*

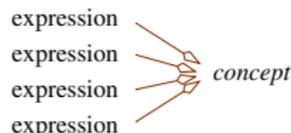


Not Bijective: Synonymy

Synonymy (or paraphrase): a unique notion is denoted by multiple distinct expressions.

[Does perfect synonymy exist?]

- *ocular prosthesis* (CDT, HCCT)
artificial eye (SNOMED CT, AOD)
- *myocardial infarction* (AOD, NDFRT, MEDLINEPLUS, NCI, ...)
myocardial infarct (MeSH)
heart attack (CDP, SNOMED CT, WHOART, MedDRA, AOD, ICPC2P...)
- *spondylarthropathy* (MeSH, CSP)
spondyloarthropathy (MedDRA, SNOMED CT, RCD)
spondarthropathy (RCD)
Disorder of joint of spine (SNOMED CT)
Marie-Strumpell spondylitis (OMIM)...



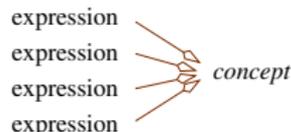


Not Bijective: Synonymy

Synonymy (or paraphrase): a unique notion is denoted by multiple distinct expressions.

[Does perfect synonymy exist?]

- *ocular prosthesis* (CDT, HCDT)
artificial eye (SNOMED CT, AOD)
- *myocardial infarction* (AOD, NDFRT, MEDLINEPLUS, NCI, ...)
myocardial infarct (MeSH)
heart attack (CDP, SNOMED CT, WHOART, MedDRA, AOD, ICPC2P...)
- *spondylarthropathy* (MeSH, CSP)
spondyloarthropathy (MedDRA, SNOMED CT, RCD)
spondarthropathy (RCD)
Disorder of joint of spine (SNOMED CT)
Marie-Strumpell spondylitis (OMIM)...





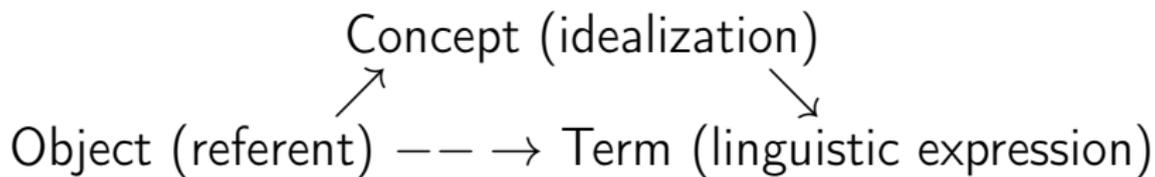
Directions

Medical Terminologies

Propositions for *normalized representations* of medical information

- 1 Medical Terminologies
 - Medical Terminologies: Why They Are Needed
 - Terminologies: General Properties
- 2 Some Terminologies in Use in the Medical Domain
 - MeSH, ICD
 - SNOMED
 - WHO-ART, MedDRA, LOINC, Procedures
- 3 Terminology Unification: the UMLS Metathesaurus
 - Metathesaurus
 - Semantic Network
 - UMLS “Specialist” Tools
- 4 Conclusion

The Semiotic Triangle



Concept

A concept:

- identifies a notion;
- is generally expressed by a **term**: its linguistic expression (e.g., *hepatomegaly*)
- generally has a definition;
genus proximus + differentia specifica
hepatomegaly: enlargement of liver
- generally has a **unique identifier** (e.g., C0019209).

Hepatomegalija: abnormalno povećanje jetre



Ambiguity and Synonymy

Normative terminology → no ambiguity.

- The **preferred term** plays the role of a norm; synonymous terms may be explicitly specified
- Reduction of polysemy and imprecision by the use of **multiword terms** where appropriate
nasal sinus vs skin sinus
myocardial infarction vs pulmonary infarction

Hierarchy

A concept may be *more specific* than another

- *heart disease* (generic)
 - *myocardial infarction*
 - *acute myocardial infarction* (specific)

A concept may be a *part of* another

- *aorta* (whole)
 - *aortic arch* (part)

Hierarchy: Multiple Classification

A unique concept is the *child of multiple fathers*

- Bacterial Infections and Mycoses
 - Infection
 - Suppuration
Abscess
Lung Abscess
- Respiratory Tract Diseases
 - Lung Diseases
 - *Lung Abscess*
 - Respiratory Tract Infections
 - *Lung Abscess*

Domain: Monoaxial vs Multiaxial Terminology

Monoaxial Terminology: describes a single type of notion;

For instance, *diagnosis*

Multiaxial Terminology: describes multiple types of notions, in distinct axes;

For instance, *anatomy, etiology, diagnosis, procedures*

→ possible *post-coordination*

- 1 Medical Terminologies
 - Medical Terminologies: Why They Are Needed
 - Terminologies: General Properties
- 2 Some Terminologies in Use in the Medical Domain
 - MeSH, ICD
 - SNOMED
 - WHO-ART, MedDRA, LOINC, Procedures
- 3 Terminology Unification: the UMLS Metathesaurus
 - Metathesaurus
 - Semantic Network
 - UMLS “Specialist” Tools
- 4 Conclusion

Bibliographic Information Retrieval: The MeSH Thesaurus

Use of Medical Subject Headings (US NLM)

- Indexing of biomedical literature (US National Library of Medicine)
 - Scientific articles in the [Medline](#) database +PubMed, PubMed Central (PMC)
 - About 5600 journals in 40 languages (2015)
 - Over 26 million citations (2017)
- Health-related documents on the Web
 - In multiple languages: [Health On the Net](#)
 - In French: [CISMeF](#) catalogue

*German, French, Portuguese, Spanish, Russian, Italian, Finnish;
Dutch, Slovak, Slovene, Swedish, Norwegian, Turkish, Chinese, Arabic, Japanese...*

Domain of MeSH: Multiaxial Thesaurus

2016 data

- 27,883 **descriptors** and preferred terms (*main headings*)
- 87,000 **synonyms** (*entry terms*)
- 232,000 supplementary concept records (**chemical terms**)

1. **Anatomy [A]**
2. **Organisms [B]**
3. **Diseases [C]**
4. **Chemicals and Drugs [D]**
5. **Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]**
6. **Psychiatry and Psychology [F]**
7. **Biological Sciences [G]**
8. **Natural Sciences [H]**
9. **Anthropology, Education, Sociology and Social Phenomena [I]**
10. **Technology, Industry, Agriculture [J]**
11. **Humanities [K]**
12. **Information Science [L]**
13. **Named Groups [M]**
14. **Health Care [N]**
15. **Publication Characteristics [V]**
16. **Geographicals [Z]**

Example Chapter in MeSH: Diseases

3. Diseases [C]

- [Bacterial Infections and Mycoses \[C01\] +](#)
- [Virus Diseases \[C02\] +](#)
- [Parasitic Diseases \[C03\] +](#)
- [Neoplasms \[C04\] +](#)
- [Musculoskeletal Diseases \[C05\] +](#)
- [Digestive System Diseases \[C06\] +](#)
- [Stomatognathic Diseases \[C07\] +](#)
- [Respiratory Tract Diseases \[C08\] +](#)
- [Otorhinolaryngologic Diseases \[C09\] +](#)
- [Nervous System Diseases \[C10\] +](#)
- [Eye Diseases \[C11\] +](#)
- [Male Urogenital Diseases \[C12\] +](#)
- [Female Urogenital Diseases and Pregnancy Complications \[C13\] +](#)
- [Cardiovascular Diseases \[C14\] +](#)
- [Hemic and Lymphatic Diseases \[C15\] +](#)
- [Congenital, Hereditary, and Neonatal Diseases and Abnormalities \[C16\] +](#)
- [Skin and Connective Tissue Diseases \[C17\] +](#)
- [Nutritional and Metabolic Diseases \[C18\] +](#)
- [Endocrine System Diseases \[C19\] +](#)
- [Immune System Diseases \[C20\] +](#)
- [Disorders of Environmental Origin \[C21\] +](#)
- [Animal Diseases \[C22\] +](#)
- [Pathological Conditions, Signs and Symptoms \[C23\] +](#)

MeSH Qualifiers

“Subheadings” <https://www.nlm.nih.gov/mesh/topscope.html>

Qualifiers, or *subheadings*

- 80 qualifiers (2017)
- A qualifier associates to a descriptor to specify the facet under which it is considered
 - liver / anatomy and histology
 - liver / diagnosis
- Each qualifier can only associate with some descriptors



Example Qualifiers

<https://www.nlm.nih.gov/mesh/subhierarchy.html>

analysis

blood

cerebrospinal fluid

isolation & purification

urine

anatomy & histology

blood supply

cytology

- ultrastructure

embryology

- abnormalities

innervation

pathology

chemistry

agonists

analogs & derivatives

antagonists & inhibitors

chemical synthesis

diagnosis

diagnostic imaging

...



MeSH Indexing

Semi-automatic routine coding in on-line portals

NLM: *MTI* (Aronson et al., 2004): combination of linguistic (MetaMap) and statistical (Related Citations) indexers, use of UMLS with restriction to MeSH thesaurus

CISMeF (Rouen): recognition of MeSH descriptors and qualifiers (linguistic patterns)

HON (Geneva): recognition of MeSH descriptors (word sequences)

See also BioASQ (<http://bioasq.org/>)

Diagnoses and Statistics:

The International Classification of Diseases

International Statistical Classification of Diseases and Related Health Problems (ICD-10, WHO)

- 12 000 classes of diagnoses
- Collect information for mortality and morbidity statistics
- Access: <http://www.who.int/classifications/apps/icd/icd10online/> (en)
<http://www.who.ch/hst/icd-10/icd-10.htm> (en, fr, de)
- Exists in 42 languages
- Routine use:
 - assisted entry of hospital diagnoses
 - semi-automated coding of causes of death

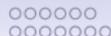
Domain of ICD-10: Monoaxial Classification

Chap.	Blocks	Title
I	A00-B99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00-E90	Endocrine , nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	H60-H95	Diseases of the ear and mastoid process
IX	I00-I99	Diseases of the circulatory system
X	J00-J99	Diseases of the respiratory system
XI	K00-K93	Diseases of the digestive system
XII	L00-L99	Diseases of the skin and subcutaneous tissue
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIV	N00-N99	Diseases of the genitourinary system
XV	O00-O99	Pregnancy , childbirth and the puerperium
XVI	P00-P96	Certain conditions originating in the perinatal period
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVIII	R00-R99	Symptoms , signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	S00-T98	Injury , poisoning and certain other consequences of external causes
XX	V01-Y98	External causes of morbidity and mortality



ICD-10 Structure: Single Hierarchy

code	label
I21	Acute myocardial infarction
I21.0	Acute transmural myocardial infarction of anterior wall
I21.1	Acute transmural myocardial infarction of inferior wall
I21.2	Acute transmural myocardial infarction of other sites
I21.3	Acute transmural myocardial infarction of unspecified site
I21.4	Acute subendocardial myocardial infarction
I21.9	Acute myocardial infarction, unspecified



Evolution : Preparation of ICD-11

- Towards a more formal structuring (cf SNOMED CT)
- Some collaboration between WHO (ICD) and IHTSDO (SNOMED)
 - Complementarity of ICD-11 and SNOMED CT
 - Linkage to improve coding and information exchange
- Computer-supported ICD revision process
 - Ontology editor (Web Protégé)
 - Collaborative editing (Wiki-style)

Properties in ICD-11

Any Category in ICD is represented by:

1. ICD Concept Title: Name of disease, disorder, or syndrome
2. Hierarchy, Type and Use
 - 2.1. Parents
 - 2.2. Type
 - 2.3. Use
3. Textual Definition(s)
4. Terms
 - 4.1. Base Index Terms
 - 4.1.1. Synonyms
 - 4.1.2. Narrower Terms
 - 4.2. Inclusion Terms
 - 4.3. Exclusion Terms
 - 4.4. Fully specified Name
5. Clinical Description
 - 5.1. Body System(s)
 - 5.2. Body Part(s) [Anatomical Site(s)]
6. Manifestation Properties
 - 6.1. Signs & Symptoms
 - 6.2. Findings
7. Causal Properties
 - 7.1. Etiology Type
 - 7.2. Agents
 - 7.3. Mechanisms
 - 7.4. Injury
 - 7.5. Risk Factors
 - 7.5.1. Immediate
 - 7.5.2. Proximal
 - 7.5.3. Distal
 - 7.6. Genomic Characteristics
8. Temporal Properties
 - 8.1. Age of Occurrence & Occurrence Frequency
 - 8.2. Development Course
9. Severity Properties
 - Option 1: No Severity subclassing
 - Option 2: Default subclassing with definitions for MILD, MODERATE, SEVERE for the disease
 - Option 3: Custom scale
10. Functioning Properties
 - 10.1. Functional impact on the person
 - 10.2. Contextual factors
 - 10.3. Body functions
11. Specific Condition Properties
12. Treatment
13. Diagnostic Criteria

Maintenance attributes

- A. Unique identifier
- B. Mapping relationships
Linkages to other systems like SNOMED etc.
- C. Other rules

ICD Coding

Numerous works on coding from free text

CCMC Challenge 2007: Cincinnati Medical Center challenge on automated ICD-9-CM coding

- 50 participants; 45 codes; top F-measure = 0.89
- Machine learning vs “symbolic” systems; negation, hypernyms, UMLS...

CLEF eHealth 2016–2018: ICD-10 coding of death certificates

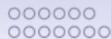
- French, 3,200 codes, F=0.83/0.87 (2017)
- French, 3,700 codes, F=0.71/0.84 (2018)
- English, 1,200 codes, F=0.85
- Hungarian, 3,200 codes, F=0.96
- Italian, 1,400 codes, F=0.95
- <https://sites.google.com/site/clefehealth2016/task-2>
- <https://sites.google.com/site/clefehealth2017/task-1>
- <https://sites.google.com/view/clef-ehealth-2018/>



Patient Care: The SNOMED Nomenclature

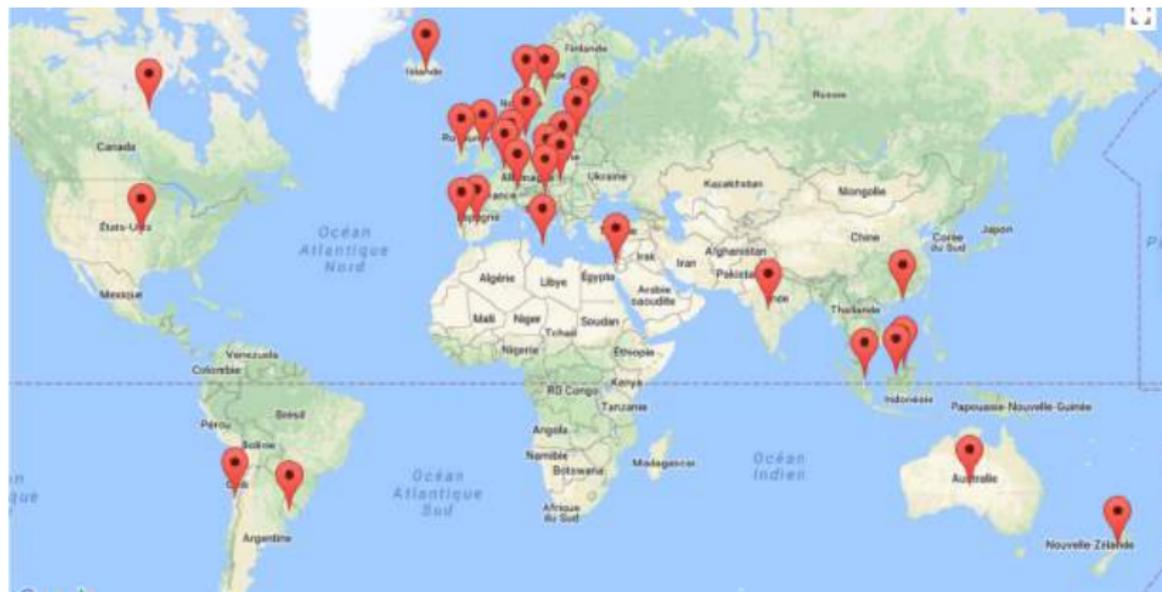
Systematized Nomenclature of Medicine

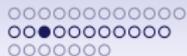
- College of American Pathologists (CAP) then IHTSDO then SNOMED International
- Very large:
 - 110,000 concepts (V3.5, 1996)
 - 360,000 active concepts (SNOMED CT, Nov 2016)
 - 1 M distinct terms (SNOMED CT, Nov 2016)
 - 1.1M distinct strings
- Usage
 - Global license in SNOMED International member states
 - Aimed at covering the needs of patient records



Member States

2017, world





Member States

2017, Europe





Domain of SNOMED: a Multiaxial Nomenclature

SNOMED v3.5: 8 orthogonal axes

Axis	Axis name	French		English	
		Terms	Concepts	Terms	Concepts
A	Artefacts, physical Activities	1600	1327	1686	1346
C	Chemical products	373	327	15940	12529
D	Diagnoses	39916	24900	42492	25193
F	[dys]Functions	19525	12874	20687	12647
G	qualifiers and relational terms	1197	891	1595	1048
J	Job	3331	1921	2303	1921
L	Living beings	480	413	26325	17678
M	Morphology	8118	4237	6171	3007
P	Procedures	197	159	31980	22156
S	Social context	1128	856	1110	858
T	Topography	13586	10258	13528	10278
X	Brand names?			363	363

according to UMLS 2006AA

SNOMED 3.5: Post-Coordination (Composition)

Allows the user to combine multiple axes to describe an observation

- Model of pathological state which combines multiple facets:

acute appendicitis	→	appendix	+	acute inflammation
		T-59200		M-41000

- Each diagnosis (D module) or procedure (P module) includes a (partial) decomposition according to these facets

→ *Increases expressive power*

SNOMED: Evolution

- The origins
 - SNOP 1965, SNOMED 1974, SNOMED II 1979
 - SNOMED v3 “International” (1993; v3.5, 1998)
- Translations (Microglossary or Full)
 - Chinese, Czech, Danish, French, German, Greek, Hungarian, Italian, Japanese, Portuguese, Russian, Slovakian, Spanish, Swedish, Turkish (2003)
 - French SNOMED v3.5 (2007: near-complete translation)
- SNOMED RT (*Reference Terminology*, 2000)
 - Adds formal descriptions
- **SNOMED CT** (*Clinical Terms*, 2002)
 - Merged SNOMED RT with the UK NHS Clinical Terms
 - Creation of IHTSDO (2007)
Australia, Canada (English), Denmark, Lithuania, The Netherlands, New Zealand, Sweden, United Kingdom, United States
 - Renamed as **SNOMED International** (2017)



SNOMED CT: Domain

Clinical finding

- Finding (*Swelling of arm*)
- Disease (*Pneumonia*)

Procedure (*Biopsy of lung*)

Observable entity (*Tumor stage*)

Body structure (*Structure of thyroid*)

- Morphologically abnormal structure (*Granuloma*)

Organism (*DNA virus*)

Substance (*Gastric acid*)

Pharmaceutical/biologic product (*Tamoxifen*)

Specimen (*Urine specimen*)

Qualifier value (*Right*)

Record artifact (*Death certificate*)

Physical object (*Suture needle*)

Physical force (*Friction*)

Events (*Flash flood*)

Environments/geographical locations (*Intensive care unit*)

Social context (*Organ donor*)

Situation with explicit content (*No nausea*)

Staging and scales (*Barthel index*)

Linkage concept

- Link assertion (*Has etiology*)
- Attributes (*Finding site*)

Special concept (*Inactive concept*)

SNOMED Clinical Terms Core Content for the January 2007 Release, CAP



SNOMED CT: Top Categories

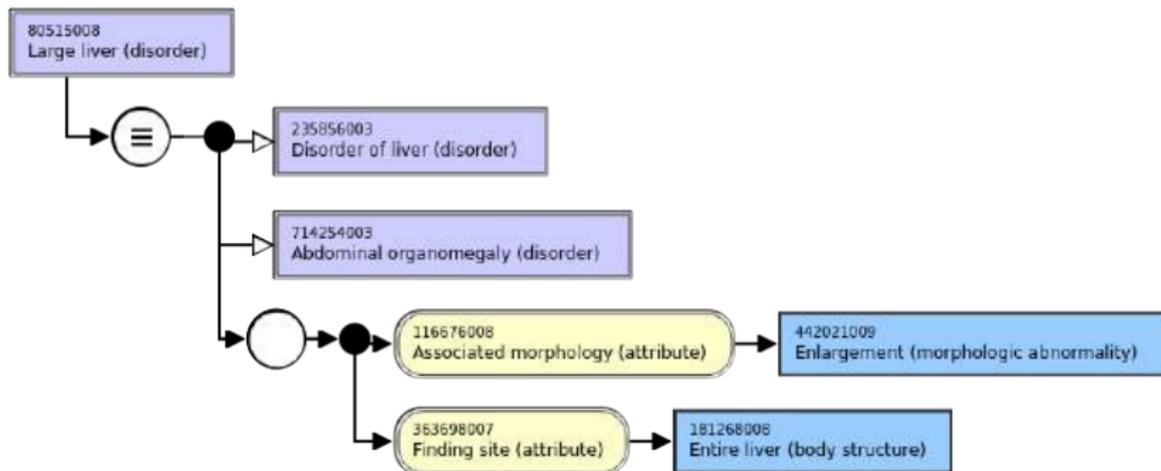
© SNOMED International 2017 v1.36.1

Inferred view ▾ Descendants Count: Off ▾

- ▼ ● SNOMED CT Concept
 - ▶ ● Body structure (body structure)
 - ▶ ● Clinical finding (finding)
 - ▶ ● Environment or geographical location (environment / location)
 - ▶ ● Event (event)
 - ▶ ● Observable entity (observable entity)
 - ▶ ● Organism (organism)
 - ▶ ● Pharmaceutical / biologic product (product)
 - ▶ ● Physical force (physical force)
 - ▶ ● Physical object (physical object)
 - ▶ ● Procedure (procedure)
 - ▶ ● Qualifier value (qualifier value)
 - ▶ ● Record artifact (record artifact)
 - ▶ ● Situation with explicit context (situation)
 - ▶ ● SNOMED CT Model Component (metadata)
 - ▶ ● Social context (social concept)
 - ▶ ● Special concept (special concept)
 - ▶ ● Specimen (specimen)
 - ▶ ● Staging and scales (staging scale)
 - ▶ ● Substance (substance)

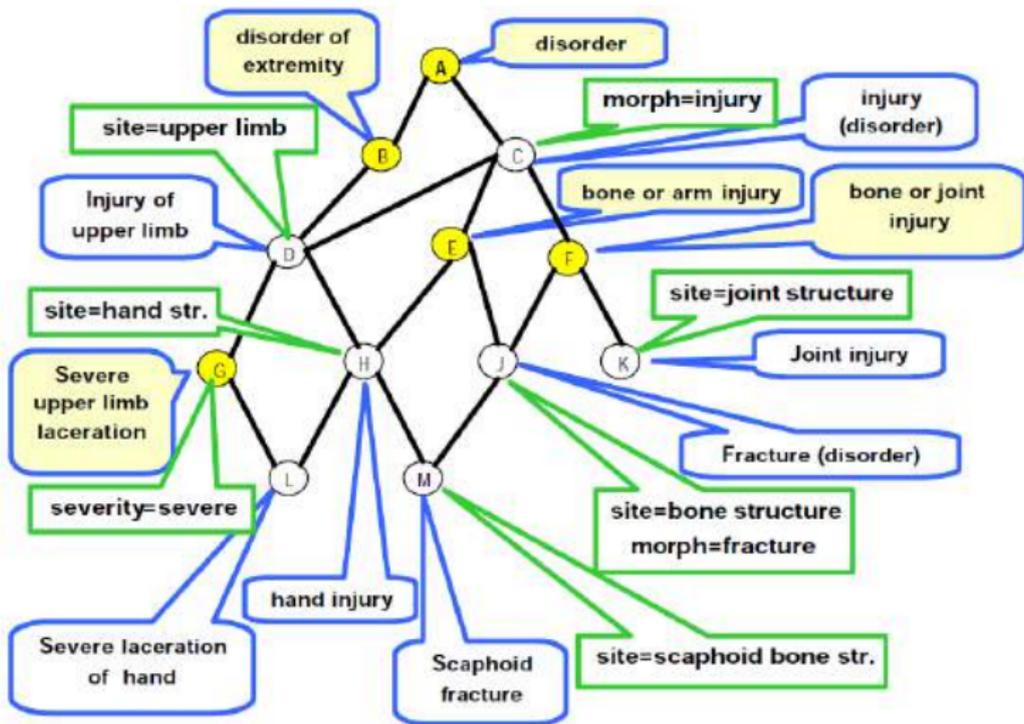
SNOMED CT: Defined Concepts

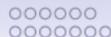
Some concepts are entirely defined based on more elementary concepts





SNOMED CT: Formal Definitions and Hierarchy





SNOMED Coding

Coding help is necessary because of the size and combinatory features of SNOMED

- *SnoCode* (Medsight Informatique, Canada): SNOMED indexing from free text reports
- Various tools



Pharmacovigilance: WHO-ART and MedDRA

Objective: Collect associations of drugs and adverse effects

- Pharmacovigilance centers
 - regional (ex: HEGP, Paris), of WHO (Uppsala)
- Need for coding adverse effects
 - **WHO-ART (WHO):** Adverse Reaction Terminology
 - **MedDRA (IFPMA):**
Medical Dictionary for Regulatory Activities
- “Signal detection”:
recurring association {drug, adverse effect}

Laboratory Test Results: LOINC

Logical Observations Identifiers Names and Codes
(LOINC Committee, Regenstrief Institute, Indianapolis)

- Types of test results
- English (US), Portuguese (Brasil), French (France, Belgium, Canada, Switzerland), Chinese (China), Italian, Russian, Spanish (Argentina, Spain, Chile), Turkish, Dutch (NL), Estonian, Korean, German (Germany, Switzerland, Austria), Italian (Switzerland), Greek (Grece)
- Direct coding within the biology laboratory



LOINC: Multiaxial Nomenclature

Six major axes:

<i>Axis</i>	<i>Description</i>
Analyte	e.g., <i>potassium</i> , <i>hemoglobin</i> , <i>hepatitis C antigen</i>
Property	e.g., <i>mass concentration</i> , <i>enzyme activity</i>
Timing	Whether the measurement is an observation at a moment of time, or an observation integrated over an extended duration of time—e.g., <i>24-hour urine</i>
Sample	The type of sample—e.g., <i>urine</i> ; <i>blood</i>
Scale	Whether the measurement is <i>quantitative</i> (a true measurement) or <i>ordinal</i> (a ranked set of options), <i>nominal</i> (e.g. <i>E. coli</i> ; <i>Staphylococcus aureus</i>), or <i>narrative</i> (e.g. <i>dictation results from x-rays</i>)
Method	Where relevant, the <i>method</i> used to produce the result or other observation

LOINC Users' Guide, Jun 2006, Regenstrief Institute & LOINC Committee



LOINC: Examples

- measurement of gammaglobulinemia (mass concentration)
- percentage of gamma globulin in blood
- measurement of gammaglobulinuria (mass concentration)
- presence of galactosuria

GAMMA GLOBULIN:MCNC:PT:SER/PLAS:QN:ELECTROPHORESIS

Code	Analyte	Property	Timing	Sample	Scale	Method
2874-6	Gamma globulin	MCNC	PT	Serum/Plasma	QN	Electrophoresis
13983-2	Gamma globulin/protein.total	MFR	PT	Serum/Plasma	QN	Electrophoresis
9745-1	Gamma globulin	MCNC	PT	Urine	QN	Electrophoresis
2309-3	Galactose	ACNC	PT	Urine	ORD	

Multiaxial description, but precoordinated coding



LOINC: Examples

- measurement of gammaglobulinemia (mass concentration)
- percentage of gamma globulin in blood
- measurement of gammaglobulinuria (mass concentration)
- presence of galactosuria

GAMMA GLOBULIN:MCNC:PT:SER/PLAS:QN:ELECTROPHORESIS

<i>Code</i>	<i>Analyte</i>	<i>Property</i>	<i>Timing</i>	<i>Sample</i>	<i>Scale</i>	<i>Method</i>
2874-6	Gamma globulin	MCNC	PT	Serum/Plasma	QN	Electrophoresis
13983-2	Gamma globulin/protein.total	MFR	PT	Serum/Plasma	QN	Electrophoresis
9745-1	Gamma globulin	MCNC	PT	Urine	QN	Electrophoresis
2309-3	Galactose	ACNC	PT	Urine	ORD	

Multiaxial description, but precoordinated coding



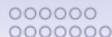
LOINC: Examples

- measurement of gammaglobulinemia (mass concentration)
- percentage of gamma globulin in blood
- measurement of gammaglobulinuria (mass concentration)
- presence of galactosuria

GAMMA GLOBULIN:MCNC:PT:SER/PLAS:QN:ELECTROPHORESIS

<i>Code</i>	<i>Analyte</i>	<i>Property</i>	<i>Timing</i>	<i>Sample</i>	<i>Scale</i>	<i>Method</i>
2874-6	Gamma globulin	MCNC	PT	Serum/Plasma	QN	Electrophoresis
13983-2	Gamma globulin/protein.total	MFR	PT	Serum/Plasma	QN	Electrophoresis
9745-1	Gamma globulin	MCNC	PT	Urine	QN	Electrophoresis
2309-3	Galactose	ACNC	PT	Urine	ORD	

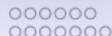
Multiaxial description, but precoordinated coding



Procedures

Varies with the country and reimbursement

- US: Common Procedural Terminology (CPT)
- France: Classification commune des actes médicaux (CCAM, Common classification of procedures)
- Procedures: directly coded when performed



Domain of the CCAM Classification

- 1 - SYSTEME NERVEUX CENTRAL, PERIPHERIQUE ET AUTONOME
- 2 - OEIL ET ANNEXES
- 3 - OREILLE
- 4 - APPAREIL CIRCULATOIRE
- 5 - SYSTEME IMMUNITAIRE ET SYSTEME HEMATOPOIETIQUE
- 6 - APPAREIL RESPIRATOIRE
- 7 - APPAREIL DIGESTIF
- 8 - APPAREIL URINAIRE ET GENITAL
- 9 - ACTES CONCERNANT LA PROCREATION, LA GROSSESSE ET LE NOUVEAU-NE
- 10 - GLANDES ENDOCRINES ET METABOLISME
- 11 - APPAREIL OSTEOARTICULAIRE ET MUSCULAIRE DE LA TETE
- 12 - APPAREIL OSTEOARTICULAIRE ET MUSCULAIRE DU COU ET DU TRONC
- 13 - APPAREIL OSTEOARTICULAIRE ET MUSCULAIRE DU MEMBRE SUPERIEUR
- 14 - APPAREIL OSTEOARTICULAIRE ET MUSCULAIRE DU MEMBRE INFERIEUR
- 15 - APPAREIL OSTEOARTICULAIRE ET MUSCULAIRE, SANS PRECISION TOPOGRAPHIQUE
- 16 - SYSTEME TEGUMENTAIRE - GLANDE MAMMAIRE
- 17 - ACTES SANS PRECISION TOPOGRAPHIQUE
- 18 - ANESTHESIES COMPLEMENTAIRES ET GESTES COMPLEMENTAIRES
- 19 - ADAPTATIONS POUR LA CCAM TRANSITOIRE

Example CCAM Chapter: Respiratory System

4 - APPAREIL CIRCULATOIRE

4.1 - ACTES DIAGNOSTIQUES SUR L'APPAREIL CIRCULATOIRE

4.1.1 - Explorations électrophysiologiques de l'appareil circulatoire

4.1.1.1 - Électrocardiographie [ECG]

4.1.1.2 - Surveillance continue de l'électrocardiogramme

4.1.1.3 - Autres explorations électrophysiologiques cardiaques

4.1.2 - Étude des pressions et des débits de l'appareil circulatoire

4.1.2.1 - Mesure des pressions et débits du cœur et des vaisseaux supracardiaques

4.1.2.2 - Mesure des pressions intravasculaires périphériques par méthode non effractive

4.1.2.3 - Mesure des pressions intravasculaires périphériques par voie vasculaire

4.1.2.4 - Surveillance continue des pressions et débits intravasculaires

4.1.3 - Échographie de l'appareil circulatoire

4.1.3.1 - Échographie du cœur et des vaisseaux intrathoraciques [Échocardiographie]

4.1.3.2 - Échocardiographie de stress

4.1.3.3 - Échographie des artères cervicocéphaliques [artères de la tête et du cou]

4.1.3.4 - Échographie des artères du membre supérieur

4.1.3.5 - Échographie de l'aorte abdominale

4.1.3.6 - Échographie des artères du membre inférieur

4.1.3.7 - Échographie des veines

4.1.3.8 - Autres échographies de l'appareil circulatoire

4.1.4 - Radiographie de l'appareil circulatoire

4.1.4.1 - Artériographie coronaire [Coronarographie]

4.1.4.2 - Artériographie de l'aorte [Aortographie]

- 1 Medical Terminologies
 - Medical Terminologies: Why They Are Needed
 - Terminologies: General Properties
- 2 Some Terminologies in Use in the Medical Domain
 - MeSH, ICD
 - SNOMED
 - WHO-ART, MedDRA, LOINC, Procedures
- 3 Terminology Unification: the UMLS Metathesaurus
 - Metathesaurus
 - Semantic Network
 - UMLS “Specialist” Tools
- 4 Conclusion



A Need for Interoperability: UMLS

Facilitate search and integration of information from multiple electronic sources of biomedical information

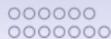
Method: Provide a bridge between existing biomedical terminologies

UMLS Unified Medical Language System

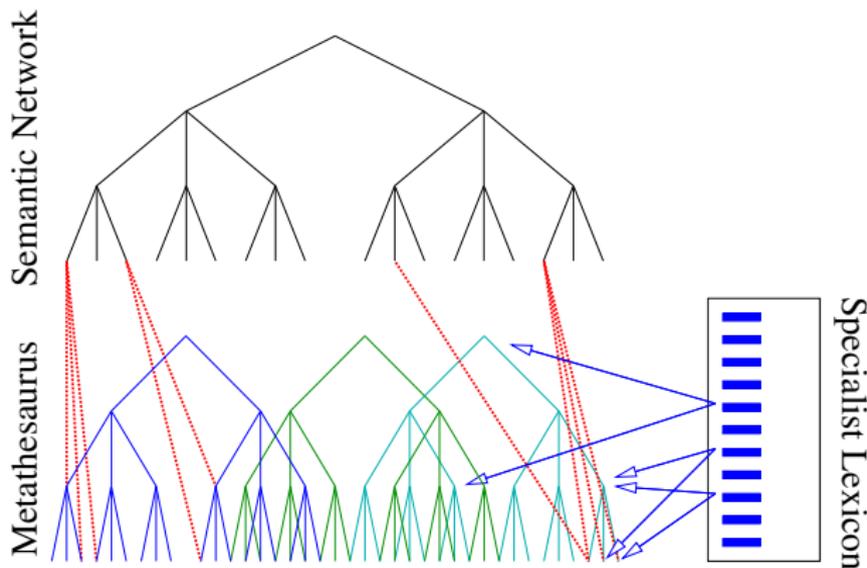
- National Library of Medicine
- Freely distributed resource (but observe rights restrictions)

<https://www.nlm.nih.gov/research/umls/>

<https://uts.nlm.nih.gov/>



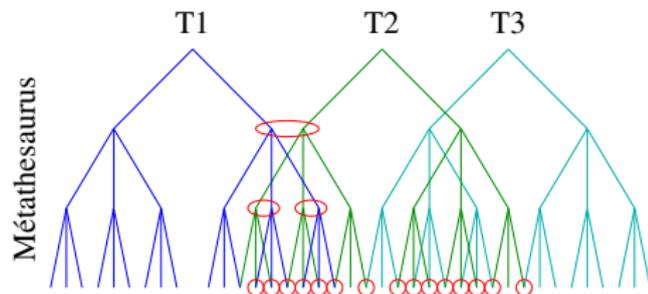
UMLS: Unifying System for Medical Terminologies



Bottom-up Approach: the “Metathesaurus”

Systematized union of two hundred biomedical terminologies

- Inventory of concepts
- Starting point = diverse terminological resources
- Identification of common concepts
- Sharing of relations
- Risks of inconsistency (attested)



2016AB: 3.4 million concepts,
10 million distinct terms

Structure of the Metathesaurus

<i>Concept</i> (CUI)	<i>Term</i> (LUI)	<i>String</i> (SUI)	
C0004238	L0004238	S0016668	
	preferred term	<i>Atrial Fibrillation</i>	preferred form
		S0016669	
		<i>Atrial Fibrillations</i>	plural variant
	L0004327	S0016899	
	synonym	<i>Auricular Fibrillation</i>	preferred form
		S0016900	
		<i>Auricular Fibrillations</i>	plural variant

UMLS: Terms and Strings

Terme	Libellé (Variante)
L0027051	<i>Myocardial Infarction</i> (PF)
L0027051	MYOCARDIAL INFARCTION (VC)
L0027051	<i>Myocardial infarction</i> (VC)
L0027051	<i>myocardial infarction</i> (VC)
L0027051	<i>Myocardial Infarctions</i> (VP)
L0027051	<i>Infarctions (Myocardial)</i> (VWP)
L0027051	<i>Infarctions, Myocardial</i> (VWP)
L0027051	<i>Infarction, myocardial</i> (VCW)
L0027051	<i>Infarction;myocardial</i> (VCW)
L0027051	<i>Infarction, Myocardial</i> (VW)
L0027051	<i>Myocardial infarction, NOS</i> (VO)
L0284112	<i>Heart attack, NOS</i> (PF)
L0284112	HEART ATTACK (VO)
L0284112	<i>Heart Attacks</i> (VO)
L0284112	<i>Heart attack</i> (VO)
L0284112	<i>heart attack</i> (VO)
L0284112	ATTACK HEART (NOS) (VCW)
L0284112	<i>Attack heart (NOS)</i> (VCW)
L0306107	<i>Infarction of heart, NOS</i> (PF)
L0306107	HEART INFARCTION (VO)
L0306107	<i>Infarction, heart</i> (VO)
L0306107	<i>Infarction;heart</i> (VO)
L0379717	<i>Cardiac infarction, NOS</i> (PF)

Terme	Libellé (Variante)
L0308108	<i>Myocardial Infarct</i> (PF)
L0308108	MYOCARDIAL INFARCT (VC)
L0308108	<i>Myocardial infarct</i> (VC)
L0308108	<i>Myocardial Infarcts</i> (VP)
L0308108	<i>Infarct, Myocardial</i> (VW)
L0308108	<i>Infarcts, Myocardial</i> (VWP)
L0308108	INFARCT MYOCARDIAL (VCW)
L0308108	<i>Infarct myocardial</i> (VCW)
L0873038	<i>Myocardial necrosis</i> (PF)
L1007490	<i>Attack coronary</i> (PF)
L1007490	ATTACK CORONARY (VC)
L1022045	<i>Myocardial infarction syndrome</i> (PF)
L1024662	<i>Necrosis myocardium</i> (PF)
L1024662	NECROSIS MYOCARDIUM (VC)
L1024664	<i>Myocardial necrosis syndrome</i> (PF)
L1088741	<i>CT - Coronary thrombosis</i> (PF)
L0586860	<i>MI - Myocardial infarction</i> (PF)
L0026809	<i>mi <1></i> (PF)
L0026809	MI (VO)
L0026809	<i>mi</i> (VO)
L1217656	<i>AMI <3></i> (PF)
L1217656	AMI (VO)
L1374411	<i>Cardiopathy necrotic</i> (PF)



Languages in the UMLS Metathesaurus (2016AB)

language	# lines
ENG	9417453
SPA	1366172
FRE	406771
POR	340009
JPN	314810
DUT	280191
ITA	224933
GER	218829
RUS	176797
CZE	162930
HUN	100794
CHI	68684
NOR	58873

language	# lines
TUR	50336
POL	45244
KOR	38664
EST	30937
SWE	26311
FIN	25489
SCR	9686
GRE	2161
LAV	1405
DAN	723
BAQ	695
HEB	485
total	10,642,957



The UMLS Semantic Network

A unifying structure superimposed over
Metathesaurus terminologies

- Semantic types
- Semantic relations
- Relation signatures

McCray AT. An upper level ontology for the biomedical domain. Comp Funct Genom 2003; 4:80-4.

Semantic Types

133 semantic types (2013–)

- General (biomedical) concepts
- Each Metathesaurus concept is categorized with one or more semantic types
- Semantic type hierarchy: *is-a* relation

Semantic Type Hierarchy (Excerpt)

Entity

Physical Object

Organism

Plant

Alga

Fungus

Virus

Rickettsia or Chlamydia

Bacterium

Animal

Invertebrate

Vertebrate

Amphibian

Bird

Fish

Reptile

Mammal

Human

Archaeon

Anatomical Structure

Embryonic Structure

Fully Formed Anatomical Structure

Body Part, Organ, or Organ Component

Tissue

Cell

Cell Component

Gene or Genome

Anatomical Abnormality

Congenital Abnormality

Acquired Abnormality

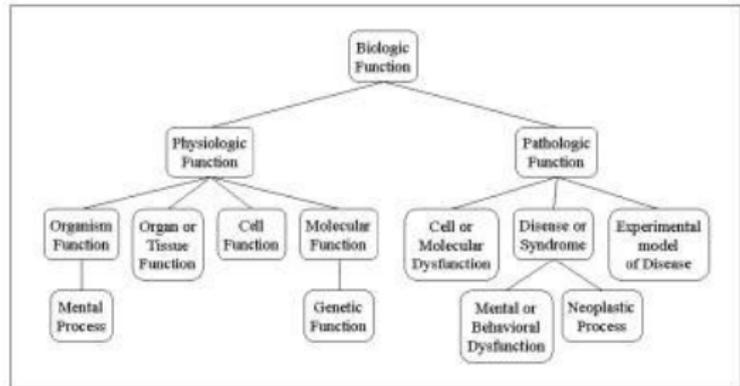
Manufactured Object

Medical Device

Drug Delivery Device

Research Device

Clinical Drug



Example Semantic Types

Acquired abnormality (STY T020, A1.2.2.2) An abnormal structure, or one that is abnormal in size or location, found in or deriving from a previously normal structure. Acquired abnormalities are distinguished from diseases even though they may result in pathological functioning (e.g., “hernias incarcerate”).

Ex. : Abscess of prostate; Hemorrhoids; Hernia, Femoral; Varicose Veins

Age Group (STY T100, A2.9.4) An individual or individuals classified according to their age.

Ex. : Adult; Infant; Premature; Adolescents; Aged, 80 and over

Alga (STY T003, A1.1.1.1) A chiefly aquatic plant that contains chlorophyll, but does not form embryos during development and lacks vascular tissue.

Ex. : Chlorella; Laminaria; Seaweed; Anabaena



Semantic Relations

54 relation types (2008AA-)

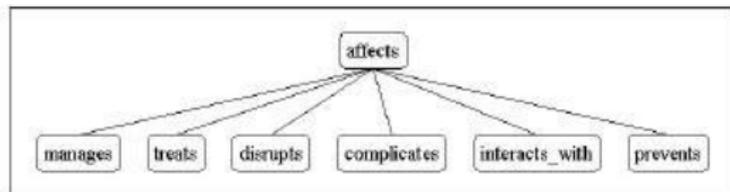
- Two concepts may be linked through a semantic relation
- Semantic relation hierarchy: *is-a* relation

Constraints on relation usage

- A given relation can only link pairs of concepts of prespecified semantic types
- $A(R)B$ means that relation R *may* exist between a concept of type A and a concept of type B

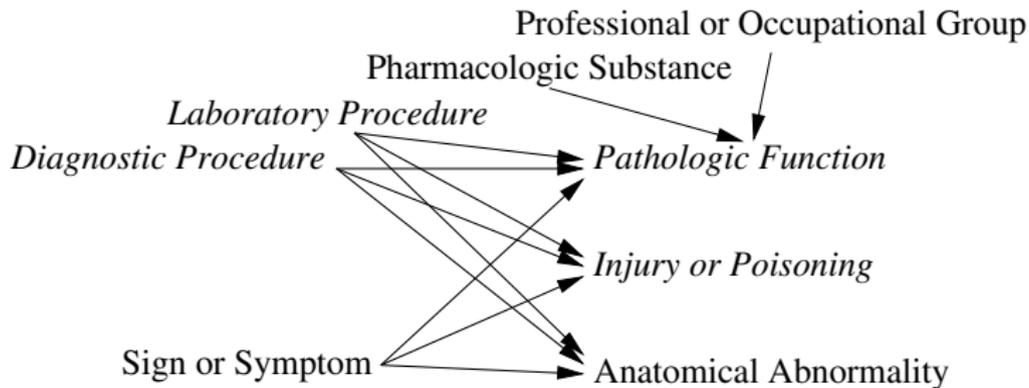
Semantic Relation Hierarchy

associated with
physically related to
part of
contains
consists of
connected to
interconnects
branch of
tributary of
ingredient of
temporally related to
co-occurs with
precedes
functionally related to
manifestation of
affects
interacts with
disrupts
prevents
complicates
manages
treats
occurs in
process of
uses
indicates
result of
brings about
produces
causes
performs
carries out
practices
exhibits
conceptually related to
property of
conceptual part of
evaluation of
measures
diagnoses
issue in
derivative of
developmental form of
degree of
measurement of
method of
analyzes
assesses effect of



Constraints on relations

Example: $A - (\text{diagnoses}) \rightarrow B$





Clustered Semantic Types: the Semantic Groups

Activities & Behaviors
 Anatomy
 Chemicals & Drugs
 Concepts & Ideas
 Devices
 Disorders
 Genes & Molecular Sequences
 Geographic Areas
 Living Beings
 Objects
 Occupations
 Organizations
 Phenomena
 Physiology
 Procedures

- 1 Medical Terminologies
 - Medical Terminologies: Why They Are Needed
 - Terminologies: General Properties
- 2 Some Terminologies in Use in the Medical Domain
 - MeSH, ICD
 - SNOMED
 - WHO-ART, MedDRA, LOINC, Procedures
- 3 Terminology Unification: the UMLS Metathesaurus
 - Metathesaurus
 - Semantic Network
 - UMLS “Specialist” Tools
- 4 Conclusion



The Specialist Lexicon

General English lexicon which includes a large number of biomedical terms

- 490 000 entries (2016AB)
- entry, lemma, POS, syntactic features (agreement, case, gender), etc.

Specialist Lexicon

<https://lsg3.nlm.nih.gov/Specialist/Home/>



Word Variants

Morphological relations

- abbreviation, orthographic variant
- semantically related terms
ocular/adj/eye/noun
- derivationally related terms
amenorrhoeic/adj/amenorrhoea/noun
- neo-classical compounds
abdomin(o)/abdomen/radical

Lexical Programs

Use Specialist Lexicon data to process biomedical terms

- Normalisation (norm)
- Segmentation into normalized “words” (wordind)
- Lexical variant generation (lvlg)

Specialist Lexical Tools

<https://lexsrv3.nlm.nih.gov/Specialist/Home/>

<https://lexsrv3.nlm.nih.gov/webapps/WebLvg.2017/jsp/lvg/getLvg.jsp>

Text Analysis Programs

Use Specialist Lexicon data to process biomedical texts

- **Word segmentation** (Tokenizer)
- **Lexicon** access (LexicalLookup)
- Part-of-Speech tagger client
- Variant generation (**derived words**) (VariantLookUp)
- **Chunking** (Parser)
- Creation of document **index** (IndexMaker)
- Creation of lexicon **index** (IndexLexicon)

Specialist Text Tools

<https://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/textTools/current/Usages/>

Controlled Indexing with the UMLS

MetaMap (Aronson, AMIA 2001)

Mapping text to UMLS Metathesaurus concepts
(and Semantic Network semantic types)

- POS tagging
- Chunking
- Variant detection
- Produces ranked list of UMLS concepts in input text

<https://mmtx.nlm.nih.gov/>

- MetaMapLite
<https://mmtx.nlm.nih.gov/MetaMapLite.shtml>

MetaMap Example

Physical fitness is a major determinant of femoral neck and lumbar spine bone mineral density.

Phrase: "Physical fitness"

- Meta Candidates (6):
 - 1000 Physical Fitness [Idea or Concept]
 - 861 Physical [Functional Concept]
 - 861 Fitness [Daily or Recreational Activity]
 - 861 Physical (Physical assessment findings) [Finding]
 - 789 Fit (Seizures) [Sign or Symptom]
 - 789 Fit (Fit and well) [Finding]
- Meta Mapping (1000):
 - 1000 Physical Fitness [Idea or Concept]

Phrase: "is"

- Meta Candidates (0): <none>
- Meta Mapping: <none>

Phrase: "a major determinant"

- Meta Candidates (1):
 - 694 Major [Qualitative Concept]
- Meta Mapping (694):

- 1 Medical Terminologies
 - Medical Terminologies: Why They Are Needed
 - Terminologies: General Properties
- 2 Some Terminologies in Use in the Medical Domain
 - MeSH, ICD
 - SNOMED
 - WHO-ART, MedDRA, LOINC, Procedures
- 3 Terminology Unification: the UMLS Metathesaurus
 - Metathesaurus
 - Semantic Network
 - UMLS “Specialist” Tools
- 4 Conclusion

Conclusion

- Terminologies are an **important resource** for Natural Language Processing in the biomedical domain
- Various **needs** led to various terminologies
- Need to **extend coverage of languages** other than English
- Meeting point of **language, knowledge representation, translation**