Co-financed by the Connecting Europe Facility of the European Union

Code systems, terminologies, and ontologies in health care

DDH - Module A - lecture 2

Enea Parimbelli - enea.parimbelli@unipv.it

Lecture summary

- What are we talking about?
- Codes for everything! (with examples)
  - ATC
  - LOINC
  - MESH
- Beyond «just codes»: Ontologies
  - SNOMED
  - GENE ONTOLOGY

Co-financed by the Connecting Europe Facility of the European Union

Semantic standardization

- "consensus" on the meaning of the data, i.e., the information collected
- Ingredients: concepts, coding of concepts, terms, and their relationships
- Fundamental to any data collected in health care:
  - Laboratory data, Medical records, Medical imaging, Molecular data (mutations, genes, proteins)

Terms, codes

- A term is associated with a medical concept
- Different terms can be used for the same concept
- A single alphanumeric code is associated with each distinct concept
- Concepts can be linked together with a series of relationships (is-a, part-of, cause, ...)

Co-financed by the Connecting Europe Facility of the European Union

Types of terminologies and coding systems

- Enumerative systems:
  - They list (in advance) all possible terms that might be used
  - They are strongly affected by the purpose for which the coding system is built
- Combinatorial systems:
  - The system attempts to construct complex terms from a set of primary or elementary terms
  - They require rules for composition that can be very complex.

Co-financed by the Connecting Europe Facility of the European Union

Terminology and coding systems

- Coding and classification systems:
  - ICD-x (9,10,11, ...), ICD-9-CM, DRG, ATC (enumerative), LOINC
- Coding systems and controlled terminology:
  - SNOMED-CT (Sistematized NOmenclature in MEDicine) International Health Terminology Standard Organization
- Meta-thesaurus:
  - UMLS (Unified Medical Language System)

Co-financed by the Connecting Europe Facility of the European Union

Some coding systems, terminologies, and ontologies

- ATC
- LOINC
- MESH
- SNOMED
- ICD-x (9-10-11)
- UMLS
- GENE ONTOLOGY

Co-financed by the Connecting Europe Facility of the European Union

The ATC / DDD system

- 1969: First drug coding activities by the Drug Utilisation Research Group (DURG), experts from the WHO
  - Activities to modify and extend the European Pharmaceutical Market Research Association (EPhMRA) classification system
- 1976: A group of Norwegian researchers published the system known as ATC: Anatomical therapeutic chemical classification system.
  - Also associated is the technical unit DDD: Defined Daily Dose "the average daily dose taken per day for the main indication in an adult.″
- 1982: ATC is maintained by the WHO Collaborating Centre for Drug Statistics Methodology in Oslo.

Co-financed by the Connecting Europe Facility of the European Union

Structure of the ATC system

- The ATC classification uses 5 levels to categorize substances, in accordance with the organ or system on which they act, and on their chemical, pharmaceutical, and therapeutic properties.
- There are 14 major groups (1st level), with a pharmacological/therapeutic subgroup (2nd level). The 3rd and 4th levels are chemical/pharmacological/therapeutic subgroups, and the 5th level represents the chemical.

Example - metformin

Level	Code	Content	
1	A	Alimentary tract and metabolism	Anatomical main group
2	A10	Drugs used in diabetes	Therapeutic subgroup
3	A10B	Oral blood glucose lowering drugs	Pharmacological subgroup
4	A10BA	Biguanides	Chemical subgroup
5	A10BA02	Metformin	Chemical substance

Co-financed by the Connecting Europe Facility of the European Union

LOINC

- The Logical Observation Identifiers Names and Codes (LOINC) database provides a set of names and codes to identify clinical and laboratory tests
- The purpose is to facilitate the exchange and collection of information, such as serum hemoglobin, potassium, or vital signs: *identification of test results or clinical observations*
- Widely used in LABS!

Co-financed by the Connecting Europe Facility of the European Union

A LOINC record

- Unique numeric code
- Component - e.g., potassium, hemoglobin
- Measured properties - e.g., concentration, enzyme activity
- Time - whether the measurement is instantaneous or integrated over 24 hours
- Type of sample - e.g., urine; blood.
- Scale type - e.g., continuous, ordinal, nominal, narrative
- Measurement method

Co-financed by the Connecting Europe Facility of the European Union

An example

LOINC code LOINC name (component:property:timing:specimen:scale:method)

LOINC_NUM	COMPONENT	PROPERTY	TIME_ASPCT	SYSTEM	SCALE_TYP	METHOD_TYP
10000-8	R WAVE DURATION.LEAD AVR	TIME	PT	HEART	QN	EKG
10001-6	R WAVE DURATION.LEAD I	TIME	PT	HEART	QN	EKG
10002-4	R WAVE DURATION.LEAD II	TIME	PT	HEART	QN	EKG
10003-2	R WAVE DURATION.LEAD III	TIME	PT	HEART	QN	EKG
10004-0	R WAVE DURATION.LEAD V1	TIME	PT	HEART	QN	EKG
10005-7	R WAVE DURATION.LEAD V2	TIME	PT	HEART	QN	EKG
10006-5	R WAVE DURATION.LEAD V3	TIME	PT	HEART	QN	EKG
10007-3	R WAVE DURATION.LEAD V4	TIME	PT	HEART	QN	EKG
10008-1	R WAVE DURATION.LEAD V5	TIME	PT	HEART	QN	EKG
10009-9	R WAVE DURATION.LEAD V6	TIME	PT	HEART	QN	EKG

Co-financed by the Connecting Europe Facility of the European Union

Other terminologies: MeSH

- MeSH: Medical Subject Headings
- Developed by NLM
- Index medical literature
  - Medline
- Terms are in hierarchies and appear in multiple positions in different hierarchies
- Site: http://www.ncbi.nih.gov - Pubmed searches

Ontologies

"An ontology is an explicit specification of a conceptualization"

Gruber. **A Translation Approach to Portable Ontology Specification.** ***Knowledge Acquisition*****. Vol. 5 1993. 199-220**

Co-financed by the Connecting Europe Facility of the European Union

Ontologies

"An ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base"

Bernaras, Laresgoiti, Correra. **Building and Reusing Ontologies for Electrical Network Applications. ECAI96. 12th European Conference on Artificial Intelligence. Ed. John Wiley & Sons. 298-302**

Co-financed by the Connecting Europe Facility of the European Union

Ontologies

- Ontologies are representations of concepts
- Ontologies are thus models, an abstract view of the world we need to represent.
- Ontologies enable the construction of structured vocabularies and provide useful definitions and mechanisms for "reasoning" about the vocabularies themselves

Co-financed by the Connecting Europe Facility of the European Union

Concepts=entities: Universals and instances

- An entity is everything that exists, things, processes, functions and qualities, beliefs and actions, documents and software
- Entities include "universals" and "instances"
- The distinction between universal and instance is fundamental to understanding how to represent data

Co-financed by the Connecting Europe Facility of the European Union

Ontologies and representations

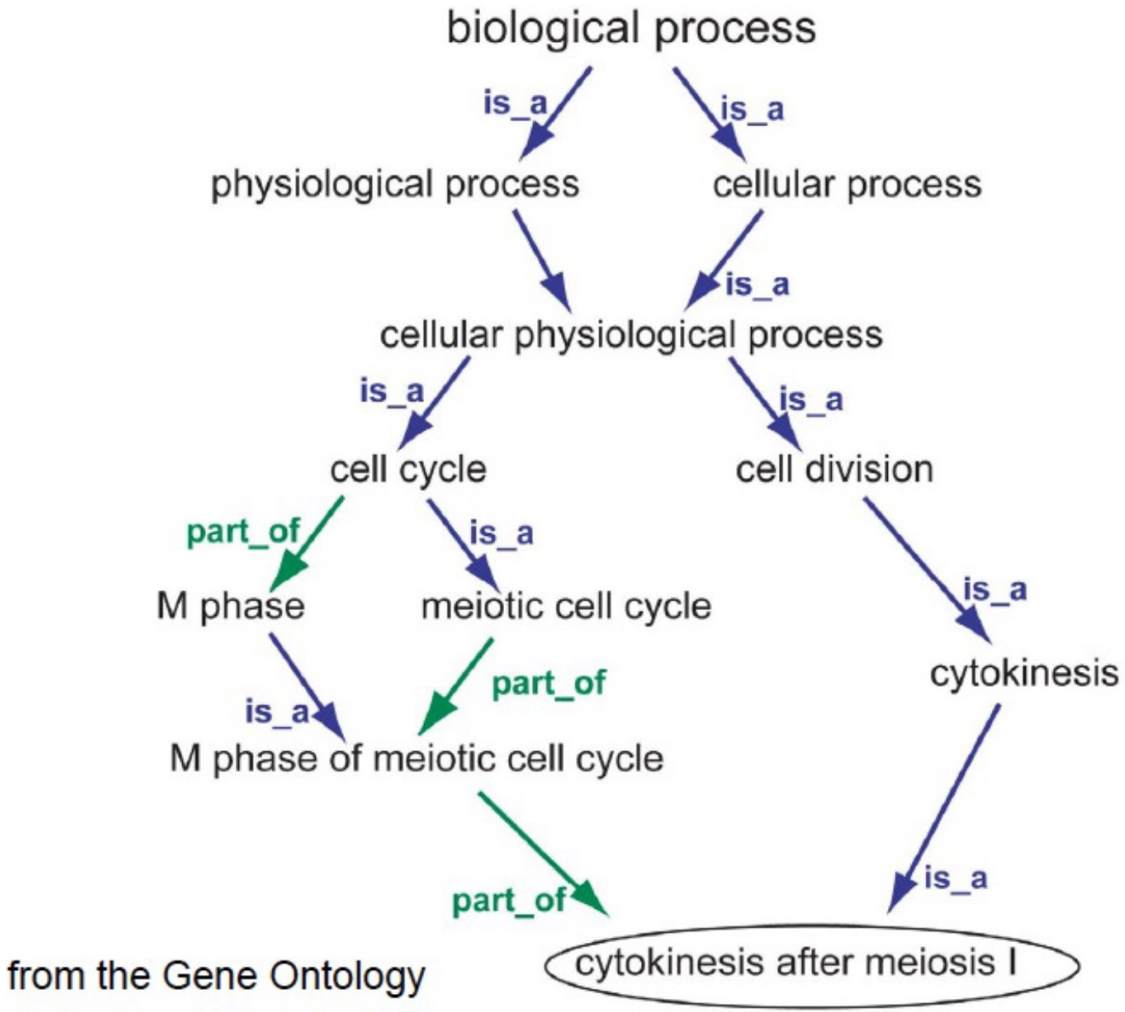
- Ontologies are usually represented as graphs
- A graph is a set of elements called nodes or vertices connected to each other by arcs or sides
- In our case, the nodes are the concepts (the universals) with their possible instances, and the arcs are the relationships between the concepts

Co-financed by the Connecting Europe Facility of the European Union

Types of relationships

- The relationships that can be expressed in ontologies are of several types:
  - *part-of*:
    - Body -> Arm -> Hand -> Fingers
  - *is-a*
    - Infection -> Hepatitis -> Viral Hepatitis -> Hepatitis A
  - *cause*
    - Plaque -> Thrombus -> Infarction -> Arrhythmia

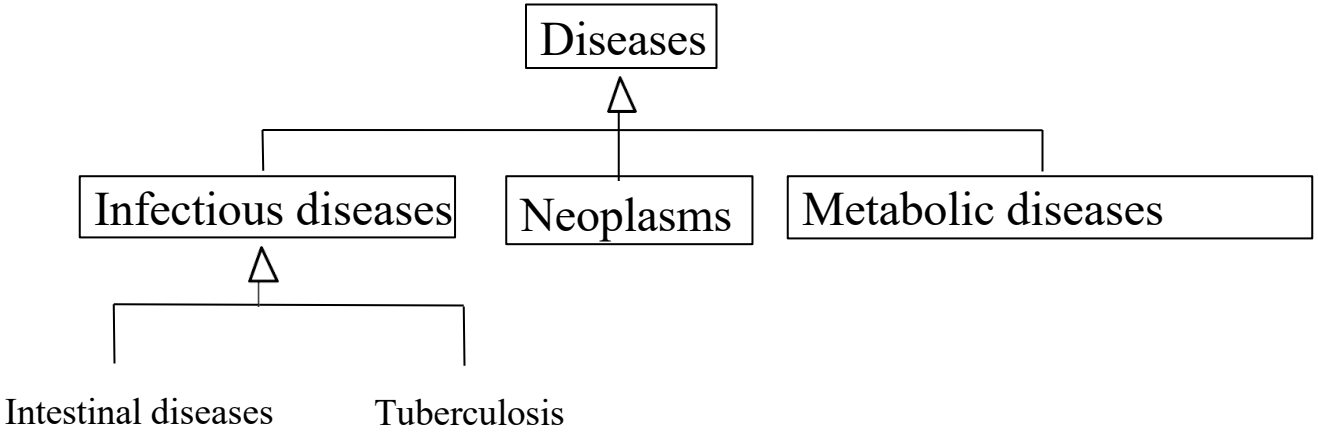
Co-financed by the Connecting Europe Facility of the European Union



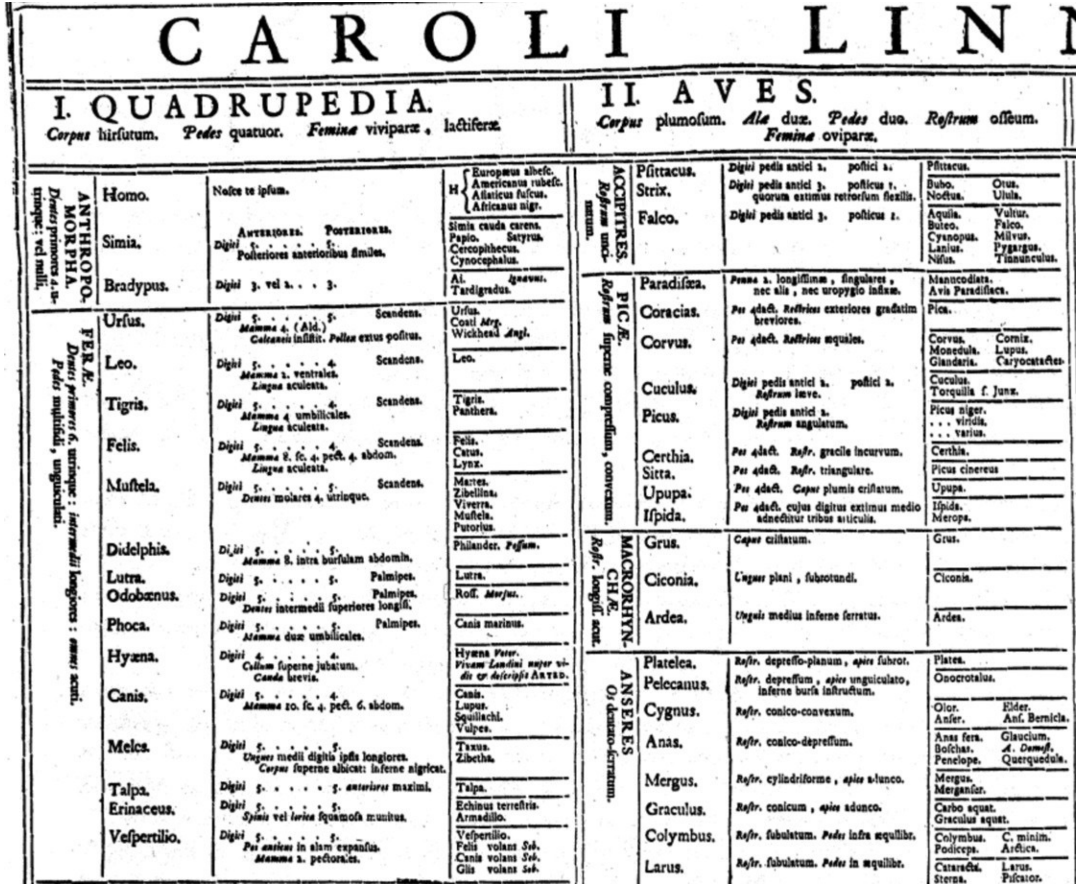
from the Gene Ontology

Classification hierarchies or Taxonomies

A structured set of ideas, organized around one of the attributes or axes.
The concepts range from the most general to the most specific



A tree of concepts: in a tree each node can have at most one parent (incoming arc). The tree has a single node with no outgoing arcs called the root of the tree. Each node that has no incoming arcs is called a leaf (leaf node). ➔ e.g. ICD



CAROLI LINN

I. QUADRUPEDIA.
Corpus hirsutum. Pedes quatuor. Feminæ viviparæ, lactiferæ.

ANTHROPOMORPHA. Dentes primores 4. utrinque; vel nulli.
Homo. Nosce te ipsum. H {Europæus albesc. Americanus rubesc. Asiaticus fuscus. Africanus nigr.
Simia. Digiti 5. 5. Anteriores. Posteriores. Posteriores anterioribus similes. Simia cauda carens. Papio. Satyrus. Cercopithecus. Cynocephalus.
Bradypus. Digiti 3. vel 2. 3. Ai. Ignavus. Tardigradus.

FERÆ. Dentes primores 6. utrinque: intermedii longiores: omnes acuti. Pedes multifidi, unguiculati.
Ursus. Digiti 5. 5. Scandens. Mammæ 4. (Abd.) Calcaneis insistit. Pollex extus positus. Ursus. Coati Marg. Wickhead Angl.
Leo. Digiti 5. 4. Scandens. Mammæ 2. ventrales. Lingua aculeata. Leo.
Tigris. Digiti 5. 4. Scandens. Mammæ 4 umbilicales. Lingua aculeata. Tigris. Panthera.
Felis. Digiti 5. 4. Scandens. Mammæ 8. sc. 4. pect. 4. abdom. Lingua aculeata. Felis. Catus. Lynx.
Mustela. Digiti 5. 5. Scandens. Dentes molares 4. utrinque. Martes. Zibellina. Viverra. Mustela. Putorius.
Didelphis. Digiti 5. 5. Mammæ 8. intra bursulam abdomin. Philander. Possum.
Lutra. Digiti 5. 5. Palmipes. Lutra.
Odobænus. Digiti 5. 5. Palmipes. Dentes intermedii superiores longiss. Ross. Morsus.
Phoca. Digiti 5. 5. Palmipes. Mammæ duæ umbilicales. Canis marinus.
Hyæna. Digiti 4. 4. Collum superne jubatum. Cauda brevis. Hyæna Veter. Vivam Londini nuper vidi & descripsi ARTED.
Canis. Digiti 5. 4. Mammæ 10. sc. 4. pect. 6. abdom. Canis. Lupus. Squillachi. Vulpes.
Meles. Digiti 5. 5. Ungues medii digitis ipsis longiores. Corpus superne albicat: inferne nigricat. Taxus. Zibetha.
Talpa. Digiti 5. 5. anteriores maximi. Talpa.
Erinaceus. Digiti 5. 5. Spinis vel lorica squamosa munitus. Echinus terrestris. Armadillo.
Vespertilio. Digiti 5. 5. Pes anticus in alam expansus. Mammæ 2. pectorales. Vespertilio. Felis volans Seb. Canis volans Seb. Glis volans Seb.

II. AVES.
Corpus plumosum. Alæ duæ. Pedes duo. Rostrum osseum. Feminæ oviparæ.

ACCIPITRES. Rostrum uncinatum.
Psittacus. Digiti pedis antici 2. postici 2. Psittacus.
Strix. Digiti pedis antici 3. posticus 1. quorum extimus retrorsum flexilis. Bubo. Otus. Noctua. Ulula.
Falco. Digiti pedis antici 3. posticus 1. Aquila. Vultur. Buteo. Falco. Cyanopus. Milvus. Lanius. Pygargus. Nisus. Tinnunculus.

PICÆ. Rostrum superne compressum, convexum.
Paradisæa. Pennæ 2. longissimæ, singulares, nec alis, nec uropygio insixæ. Manucodiata. Avis Paradisiaca.
Coracias. Pes 4dact. Rectrices exteriores gradatim breviores. Pica.
Corvus. Pes 4dact. Rectrices æquales. Corvus. Cornix. Monedula. Lupus. Glandaria. Caryocatactes.
Cuculus. Digiti pedis antici 2. postici 2. Rostrum læve. Cuculus. Torquilla s. Junx.
Picus. Digiti pedis antici 2. Rostrum angulatum. Picus niger. ... viridis. ... varius.
Certhia. Pes 4dact. Rostr. gracile incurvum. Certhia.
Sitta. Pes 4dact. Rostr. triangulare. Picus cinereus.
Upupa. Pes 4dact. Caput plumis cristatum. Upupa.
Ispida. Pes 4dact. cujus digitus extimus medio adnectitur tribus articulis. Ispida. Merops.

MACRORHYNCHÆ. Rostr. longiss. acut.
Grus. Caput cristatum. Grus.
Ciconia. Ungues plani, subrotundi. Ciconia.
Ardea. Unguis medius inferne serratus. Ardea.

ANSERES. Os dentato-serratum.
Platelea. Rostr. depresso-planum, apice subrot. Platea.
Pelecanus. Rostr. depressum, apice unguiculato, inferne bursa instructum. Onocrotalus.
Cygnus. Rostr. conico-convexum. Olor. Elder. Anser. Ans. Bernicla.
Anas. Rostr. conico-depressum. Anas fera. Glaucium. Boschas. A. Domest. Penelope. Querquedula.
Mergus. Rostr. cylindriforme, apice adunco. Mergus. Merganser.
Graculus. Rostr. conicum, apice adunco. Carbo aquat. Graculus aquat.
Colymbus. Rostr. subulatum. Pedes infra æquilibr. Colymbus. C. minim. Podiceps. Arctica.
Larus. Rostr. subulatum. Pedes in æquilibr. Cataracta. Larus. Sterna. Piscator.

Linnaeus,
18th century

Co-financed by the Connecting Europe Facility of the European Union

eXplainable Artificial Intelligence in healthcare Management
2020-EU-IA-0098

Nanochromis teugelsi Lamboj & Schelly, 2006

Upload your photos and videos
Pictures | Google image

Nanochromis teugelsi
Picture by Judy, T.

Classification / Names Common names | Synonyms | Catalog of Fishes(genus, species) | ITIS | CoL | WoRMS | Cloffa

Actinopteri (ray-finned fishes) > Cichliformes (Cichlids, convict blennies) > Cichlidae (Cichlids) > Pseudocrenilabrinae
Etymology: *Nanochromis:* Latin, nannus = small + Greek, chromis = a fish, perhaps a perch (Ref. 45335); *teugelsi:* Named after Guy G. Teugels, former curator of fishes at the Africa Museum in Tervuren, Belgium (Ref. 57723).

Environment: milieu / climate zone / depth range / distribution range Ecology

Freshwater; benthopelagic. Tropical

Distribution Countries | FAO areas | Ecosystems | Occurrences | Point map | Introductions | Faunafri

Africa: Kwilu and Kasai drainages, Lake Tumba and the main river channel between both (middle Congo River basin), in Democratic Republic of the Congo (Ref. 57723).

Size / Weight / Age

Maturity: L_m ? range ? - ? cm
Max length : 5.6 cm SL male/unsexed; (Ref. 57723)

Still used in modern taxonomy!

Lamboj, Anton, and Robert Schelly. "Nanochromis teugelsi, a new cichlid species (Teleostei: Cichlidae) from the Kasai region and central Congo basin." *Ichthyological Exploration of Freshwaters* 17.3 (2006): 247.

Co-financed by the Connecting Europe Facility of the European Union

The ingredients for defining ontologies

1. Entities: universals and instances
2. Conceptualizations (this is the ontology in its core definition)
3. Their representations (graph/tree)

- Each ontology is defined on a domain, a portion of reality of interest

Co-financed by the Connecting Europe Facility of the European Union

SNOMED CT

- Systematized NOmenclature of MEDicine Clinical Terms.
- Comprehensive terminology system based on a **multi-axial** classification system built for use in IT systems such as medical records
- It is the most extensive clinical terminology available in the world, and it is multilingual.
- The content is developed and validated by ″The International Health Terminology Standards Development Organization″ (IHTSDO)

Co-financed by the Connecting Europe Facility of the European Union

SNOMED CT: structure

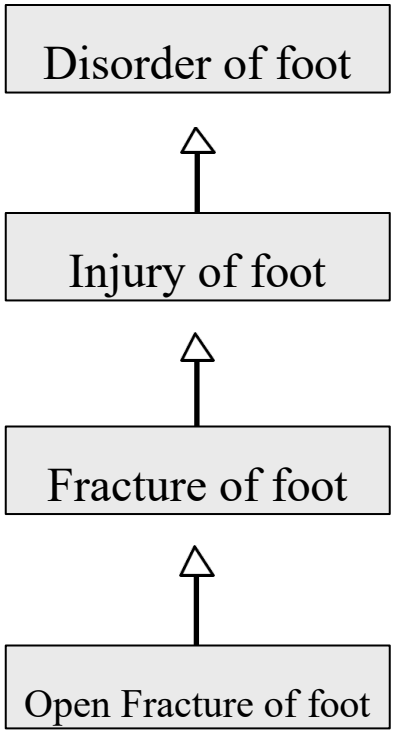
- Organized into **concepts** linked by **relationships**
- **The main relationships** are of type **IS-A** → they define **hierarchies.**
- The other relationships link the hierarchies

  - **Concepts**: clinical expressions defined by a unique numerical code called Concept ID (over 311,000 concepts)

  - **Relationships**: describe the relationships between concepts and provide their meaning

    - SNOMED consists **of 1.38 million relationships**.

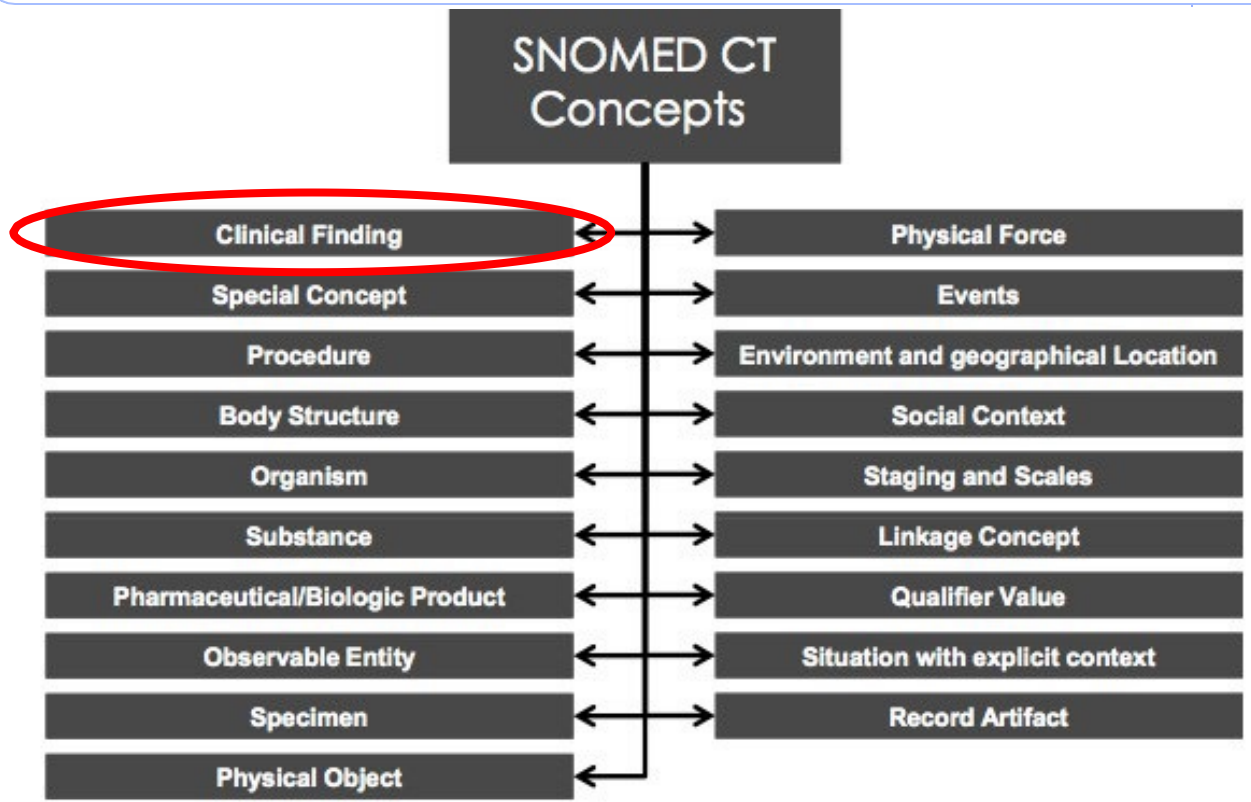
is-a relationship in SNOMED



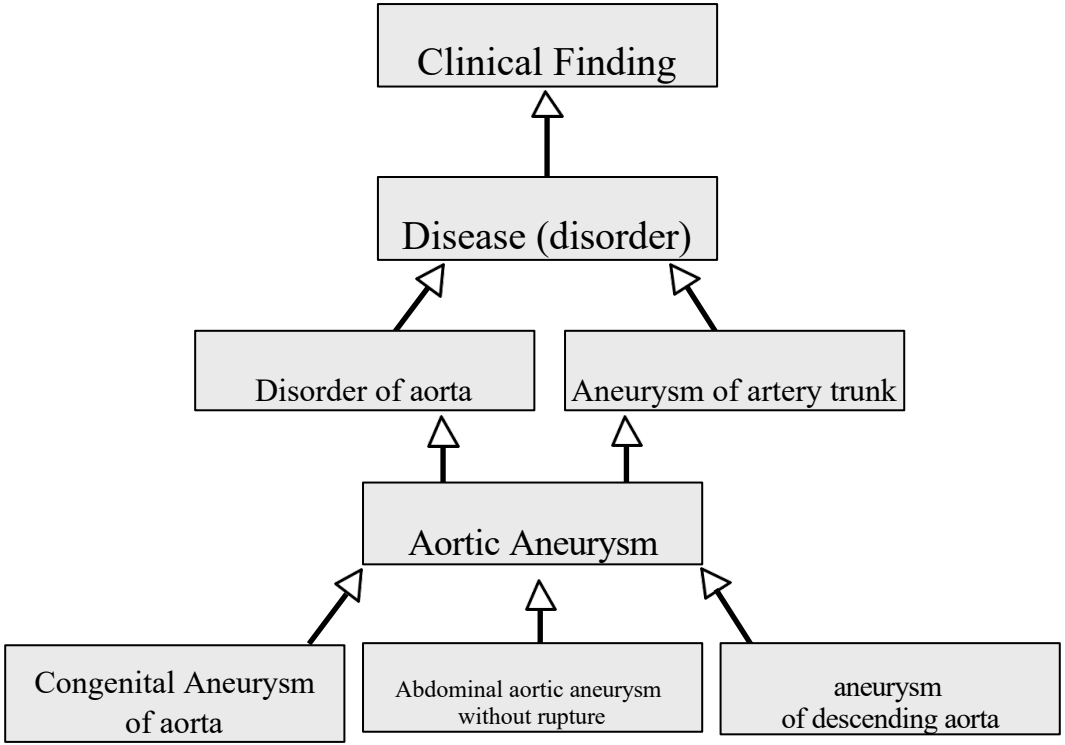
defines hierarchies from the general to the particular

Co-financed by the Connecting Europe Facility of the European Union

SNOMED CT: 19 Hierarchies



SNOMED CT clinical finding: a portion of the hierarchy



Co-financed by the Connecting Europe Facility of the European Union

Attribute relationships (qualifiers)

- An attribute relation is an association between two concepts that specifies a feature of one concept (the source of the relation) by means of another concept (the value or destination of the relation); each attribute relation has a name (the type of the relation)

pneumonia (source)

has: pathological process (relationship) = infectious process (destination)

SNOMED CT: concepts

Concepts are described by terms:

- **Preferred** term: the term of reference for the concept.
- **Synonyms**: different ways of referring to the same concept.
- **Fully Specified Name (FSN)**: a description of the concept that clarifies its meaning. FSN is unique, while preferred terms and synonyms may not be.

Co-financed by the Connecting Europe Facility of the European Union

SNOMED CT: Myocardial infarction

- ConceptID: 22298006
- Fully Specified Name: Myocardial infarction (disorder)
  DescriptionID 751689013;
- Preferred term: myocardial infarction
  DescriptionID 37436014;
- Synonym: Cardiac infarction
  DescriptionID 37442013;
- Synonym: heart attack
  DescriptionID 37443015;
- Synonym: Infarction of heart
  DescriptionID 37441018.

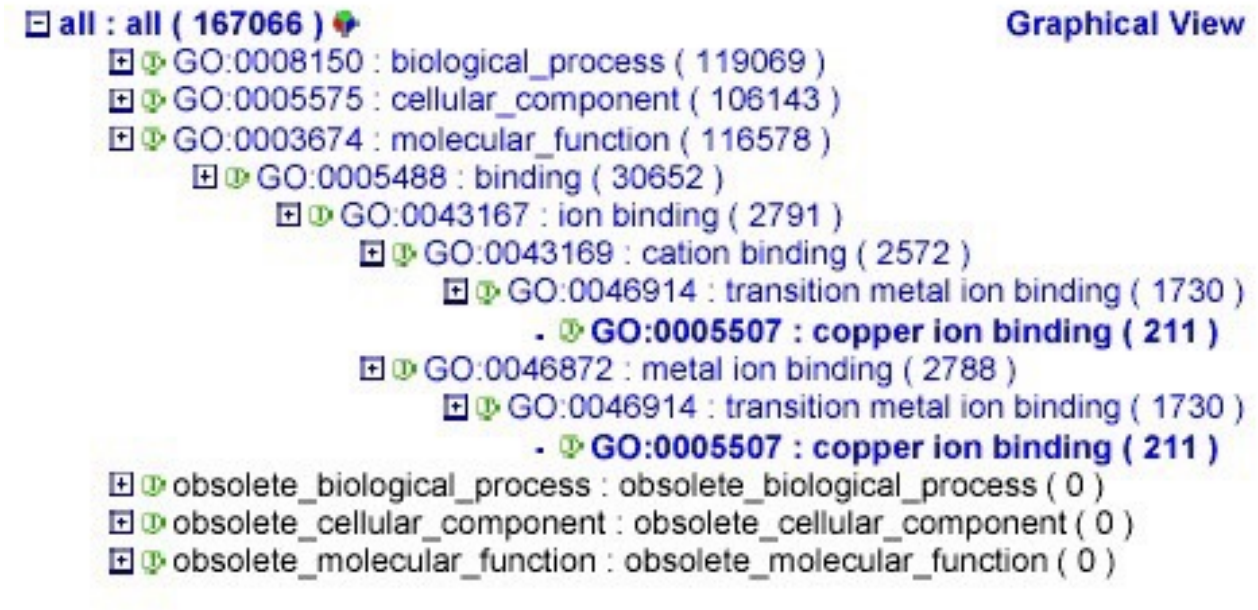
Co-financed by the Connecting Europe Facility of the European Union

Gene Ontology

- The Gene Ontology (GO) consortium is developing a controlled vocabulary for genes and gene products
- This vocabulary is used to annotate genes and proteins in databases covering data on individual organisms
- There are many groups that are members of the Consortium.
- GO terms are developed around three organizing principles :
  - Molecular function: task performed by the single gene product (e.g., transcription factor) - What type it is.
  - Biological Process: biological objective of the product (e.g. mitosis) - What it does
  - Cellular component: subcellular structures, cell sites (such as nucleus, telomere) - Where it is found
- http://www.geneontology.org/.

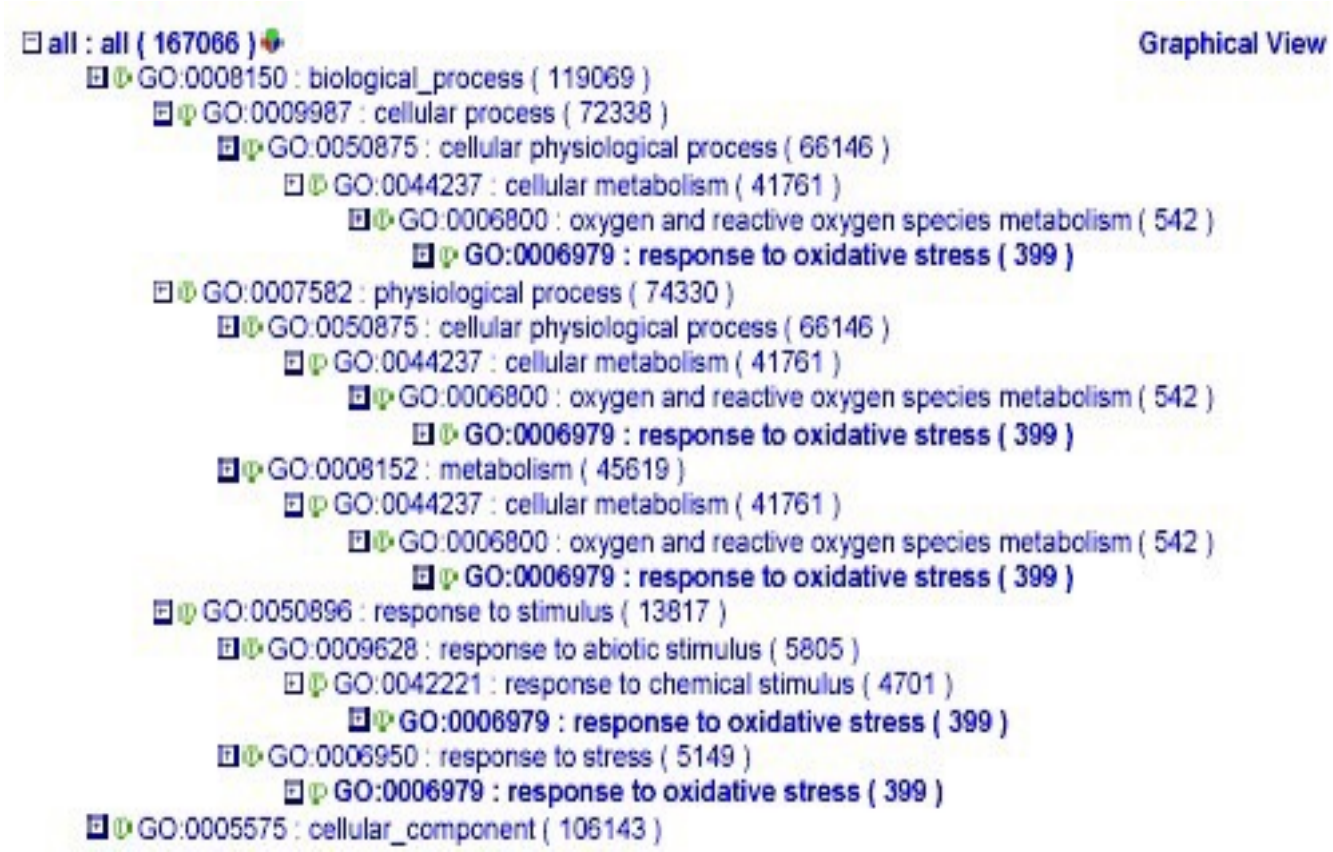
Co-financed by the Connecting Europe Facility of the European Union

Example - Molecular function



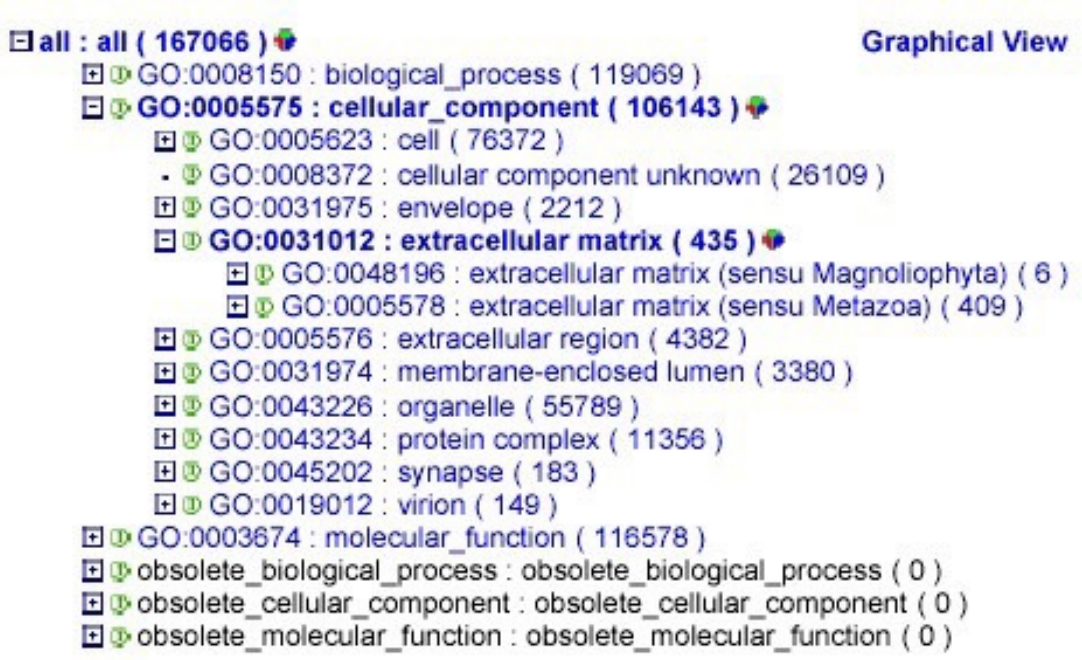
Co-financed by the Connecting Europe Facility of the European Union

Example - Biological Process



Co-financed by the Connecting Europe Facility of the European Union

Example - Cellular Component

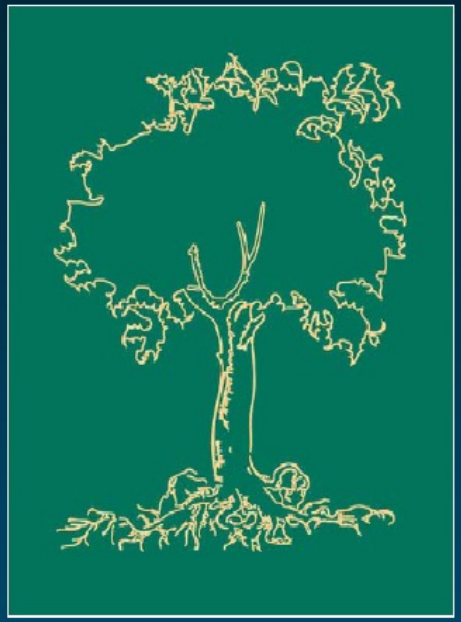


Co-financed by the Connecting Europe Facility of the European Union

Unified Medical Language System (UMLS)

- Developed to facilitate retrieval and integration of biomedical information from diverse sources
- Components (sources of knowledge)
  - Metatesaurus
  - Semantic network
  - Lexical analysis programs
- Contains more than 1 million biomedical concepts and 5 million names, which come from 100 controlled vocabularies and classification systems

- Unified
- Medical
- Language
- System



UMLS®
Unified Medical Language System®
UMLS Metathesaurus®

NLM