Chapter 2
Medical Vocabulary, Terminological Resources and Information Coding in the Health Domain

C. Duclos, A. Burgun, J.B. Lamy, P. Landais, J.M. Rodrigues, L. Soualmia, and P. Zweigenbaum

Abstract This chapter explains why it is hard to use medical language in computer applications and why the computer must adopt the human interpretation of medical words to avoid misunderstandings linked to ambiguity, homonymy and synonymy. Terminological resources are specific representations of medical language for dedicated use in particular health domains. We describe here the components of terminology (terms, concepts, relationships between concepts, definitions, constraints). The various artefacts of terminological resources (e.g. thesaurus, classification, nomenclature) are defined. We also provide examples of the dedicated use of terminological resources, such as disease coding, the indexing of biomedical publications, reasoning in decision support systems and data entry into electronic medical records. ICD 10, SNOMED CT, and MeSH are among the terminologies used in the examples. Alignment methods are described, making it possible to identify equivalent terms in different terminologies and to bridge

C. Duclos (✉)
LIM&BIO EA 3969, UFR SMBH Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny Cedex, France
e-mail: catherine.duclos@avc.aphp.fr

A. Burgun
Centre de recherche des cordeliers, 15 rue de l'école de Médecine, 75006 Paris, France

J.B. Lamy
UFR SMBU Université Paris 13, 74 rue Marcel Cachin, 93017 Bobigny Cedex, France

P. Landais
Université de Montpellier 1, 641 avenue du Doyen Gaston Giraud, 34093 Montpellier Cedex 5, France

J.M. Rodrigues
Université Jean Monnet, 10 rue Tréfilerie, 42023 Saint Etienne Cedex 2, France

L. Soualmia
Université de Rouen, Place Emile Blondel, 76821 Mont Saint Aignan Cedex, France

P. Zweigenbaum
LIMSI-CNRS, BP 133, 91403 Orsay, France

A. Venot et al. (eds.), *Medical Informatics, e-Health*, Health Informatics,
DOI 10.1007/978-2-8178-0478-1_2, © Springer-Verlag France 2014

different domains in health. We also present plans for multi-terminological servers, such as the UMLS (Unified Medical Language Systems), which provide a key vocabulary linking heterogeneous health terminologies in different languages.

Keywords Terminology • Classification • Semantics • Language • Controlled vocabulary • Coding scheme

After reading this chapter you should:

- Know the characteristics of medical language and the notions of synonymy, homonymy and ambiguity,
- Understand the requirement for the formalisation of medical language for the computerisation of health activities,
- Understand the notion of a "concept",
- Be aware of the various components used for the development of terminological systems,
- Know the definition of the different terminological systems and the issues they address,
- Be aware of the various uses of terminological systems and be able to provide examples suitable for a dedicated use,
- Be able to use terminological servers,
- Be aware of the major dedicated terminological resources in the domain of health.

2.1 Introduction

Healthcare professionals use specific health-related terminologies to express entities as diverse as diagnoses, findings, procedures, laboratory tests, drugs, anatomy, biological findings or genetics. Health terminology is complex and multifaceted.

The computerisation of health systems requires the recording and storage of large amounts of information about the health of patients and populations and expectations are high for the "intelligent" use of such information.

Humans can understand and reason from words, based on an understanding of their meaning, but computers can only compare text strings. For a computer to be able to understand medical language, resources that convey meaning are required. Terminological resources provide lists of organised concepts for specific health domains. These representations, when used to model information, can convey, contextually, the meaning of health information, enabling the computer to use this information correctly.

In this chapter, we discuss the characteristics of medical language, the need to normalise the expression of medical concepts for their use by computers and how such concepts can be represented through various artefacts (e.g. thesauri, classifications, nomenclatures), the components of which we describe here.

We then illustrate the use of these terminological resources by particular cases and present tools for viewing terminological resources (multi-terminological servers).

2.2 The Medical Vocabulary and Its Properties

2.2.1 Medical Vocabulary

Medical vocabulary has evolved with the historical development of medicine and surgery. Medical terms have been translated from Greek to Arabic, Arabic to Latin and Latin to modern languages. The coexistence of these various languages, Latin, Greek and Arabic, and of various schools of thought, such as the Aristotelian or Platonic schools, has made it difficult to develop an unequivocal single medical vocabulary. The simultaneous use of several linguistic systems has led to multiple synonyms.

2.2.2 Establishment of Medical Terms

Most medical terms are borrowed from Greek and Latin. They consist of a radical, possibly associated with a prefix or a suffix. The radical is the root of the word (for example, the radical "pharmac", from the Greek *pharmakon* refers to drugs in *pharmac*y or *pharmac*ology). A prefix is an element in front of a word, which modifies the meaning (e.g. the prefix *a*- indicates absence, as in *a*mnesia). A suffix is placed at the end of the word and also modifies its meaning (e.g. the suffix -*itis* indicates inflammation, of the larynx in laryng*itis*, of a node in aden*itis*, or a joint in arthr*itis*).

Radicals

The *kine* radical refers to movement, as in *kine*tics or a*kine*sia; here the “a” indicates an absence of movement, one of the characteristics of Parkinson’s disease.

The association of the radical *cyt*(o)- (cell) and the suffix –*logy* (study), results in “cytology”, the study of the cell. Similarly, the association of *histo*- with the suffix -*logy* gives histology, the study of tissues.

Prefixes

Absence or deprivation:	a- (*a*mnesia), *an*- (*a*naemia), ab-(*ab*stinence), *in*-(*in*somnia), *im*-(*im*maturity).
Number:	0, *nulli*- : a *nulli*para is a woman who has never given birth;
	1, *primi*-: a *primi*para has given birth once;
	n, *multi*-: a *multi*para has already given birth several times.
Quantity:	much, *poly*-: *poly*uria = much urine;
	little, *olig*-: *olig*uria, little urine, pauci- *pauci*symptomatic.
Frequency:	fast, *tachy*-: *tachy*cardia = fast heart rate;
	slow, *brady*-:*brady*cardia = slow heart rate;
	often, *pollaki*- : *pollaki*uria = needing to urinate frequently;
	rare, *spanio*-: *spanio*menorrhea = a decrease in the frequency of periods.
Site:	in the middle, *mid*renal, *meso*colon;
	in front of, *prerenal*;
	behind, *retro*caval;
	above, *supra*tentorial;
	below, *hypo*gastrium;
	next to, *para*umbilical;
	around, *peri*carditis;
	at the base, *rhiz*arthrosis;
	at the end, *acro*megaly.
Resemblance:	self; *auto*graft;
	the same, *homo*zygotic twins;
	different, *hetero*geneous, *hetero*zygotic twins.
Function:	normal, *eu*thyroidism;
	abnormal, *dys*thyroidism;
	high, *hyper*thyroidism;
	low, *hypo*thyroidism

Suffixes

- *algia* means pain, arthr*algia*, joint pain
- *osis* refers to degeneration, as in adenomat*osis*

Suffixes (continued)

- *lysis* indicates destruction: auto*lysis* for self-destruction, osteo*lysis* for bone destruction
- *ectasia* or – *cele* for dilation: bronchi*ectasis*, varico*cele*
- *sten* refers to the narrowing of the lumen of a conduit: coronary *sten*osis
- *stasis* means stagnation, chole*stasis* is bile accumulation
- *rrhoea* refers to a flow, as in rhino*rrhoea* for runny nose, or *a*meno*rrhoea*, the cessation of menstrual periods
- *oma* denotes a malignant tumor, carcin*oma*
- *tomy*, indicates an incision or opening: phlebo*tomy*, opening a vein, gastro*tomy* opening the stomach to insert a feeding tube, for example
- *stomy* is used to indicate surgical procedures in which stomata are created: colo*stomy*, creation of a stoma from the skin to the colon
- *pexy* indicates attachment: cysto*pexy,* for example, is the attachment of the bladder to the abdominal wall
- *ectomy* means removal, excision: nephr*ectomy*
- *plasty* means repair, rhino*plasty* for nose reconstruction

Some medical terms may be eponymous. In other words, they may include the name of a person (e.g. Dupuytren's disease, Hodgkin's lymphoma).

Some medical terms are acronyms, formed from the first letters of a group of words and generally pronounced letter by letter, although some create collections of letters than can be pronounced like words in their own right (BBS for Besnier, Boeck and Schaumann sarcoidosis, NSAIDs for non-steroidal anti-inflammatory drugs, and MI for myocardial infarction).

2.2.3 Properties of Medical Language: Synonymy, Polysemy, Vagueness, Ambiguity

Medical language, like any language, can be difficult to understand because of ambiguities, leading to various possible interpretations of individual words. These ambiguities may result from the definition of a given word or acronym not being universal (for example, the acronym VIP is interpreted as vasoactive intestinal peptide in gastro-enterology but as voluntary interruption of pregnancy in orthogenic departments). Alternatively, a word may be ambiguous because it has many meanings (polysemy): for example, the word "knee" may represent a joint (dislocation of the knee) or an anatomical angle (right inferior knee of the coronary artery). Polysemy may be eliminated by taking into account the context in which the term is used. Finally, the ambiguity may result from the vagueness of language: "infarction" commonly refers to the heart, but it is more precise to talk about "myocardial infarction" in this case, to differentiate between this type of infarction and mesenteric or cerebral infarction, for instance.

Medical language is also highly expressive and includes many synonyms, i.e. expressions referring to the same object (e.g., myocardial infarction, heart attack, MI).

Natural language is extremely powerful and flexible. It can deal with various degrees of precision and evolution due to changes in knowledge, and it makes it possible to understand the context even when implicit, because humans can interpret and draw inferences from a knowledge of language.

2.3 Normalising the Expression of Medical Concepts in Computing Environments to Ensure Semantic Interoperability

When information is stored in a computer system, the computer "sees" the words simply as a string of characters. The computer system can carry out logical operations on these strings (e.g., it can check whether two words are identical by comparing each character of the textual string).

The storage of information on computers is of value if it provides benefits for the healthcare provider and for the patient. For example, noting that a patient is asthmatic in his computerised medical record should be associated with an automatic reminder to vaccinate the patient against flu, because he has a chronic respiratory disorder.

This requires computerised systems to understand information and, therefore, to make use of the meaning of the information rather than its expression. This is referred to as "semantic interoperability".

Resources are required to limit the ambiguities and imprecision of natural language, and to manage synonymy. These resources must introduce elements of context reproducing the organisation of knowledge necessary for a human to interpret words.

With such resources, it is possible:

- To record clinical data and to store them in electronic patient records with the appropriate level of detail,
- To exchange clinical data between independently developed clinical information systems without human intervention and with no loss of meaning,
- To combine similar data from several independent information systems without human intervention (for example, for health monitoring systems),
- To share decision rules from clinical practice guidelines between hospitals using different, independent information systems without human intervention and to use them with the data stored in these information systems.

Use Cases

2.3.0.1 *Use Case 1*

A patient arrives in the emergency department. The reason for the consultation is entered into the computerised information system and transmitted to the attending physician. The data are imported and integrated into the patient's electronic health record. The diagnosis of acute renal colic is automatically added to the list of the patient's problems, an analgesic prescription is added to the list of treatments and the results of laboratory tests and scanner findings are added to the patient's medical history.

2.3.0.2 *Use Case 2*

A health monitoring system for detecting the exposure of the population to infectious organisms retrieves, each night, the data concerning diagnoses, symptoms and bacteriological results stored in various hospital information systems. It combines these data and analyses them, to detect the emergence of new infectious diseases.

2.4 Terminology Resource Components

An understanding of the structures and utilisation targets of the various terminology resources used in healthcare requires a definition of the lexical assumptions on which they are built. There are two approaches:

The first is based on the onomasiological theory of word formation, which gives names to a meaning, thought or concept. In this case, the different designations or terms mapping to a specific meaning are sought and qualified as synonyms; for example "necrosis of the myocardium after coronary obstruction by a thrombus" can be named "myocardial infarction" or by the acronym "MI".

The second lexicographical approach is based on the collection of different words, from which meanings or concepts are extracted (semasiological approach). In this case, the same term can have several meanings and are said to be homonyms or polysemic, as in "cold" as a level of temperature and the name of a particular illness.

2.4.1 *Triangle Concept, Term, Object or Thought, Word, Thing*

There are three main framework components based on the Ogden-Richards semiotic triangle (Ogden et al. 1923) and the modified Ogden-Richards semiotic triangle

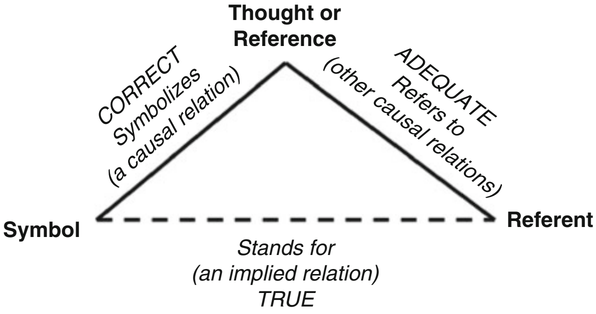

Thought or Reference
CORRECT Symbolizes (a causal relation)
ADEQUATE Refers to (other causal relations)
Symbol
Referent
Stands for (an implied relation) TRUE

Fig. 2.1 Ogden-Richards triangle

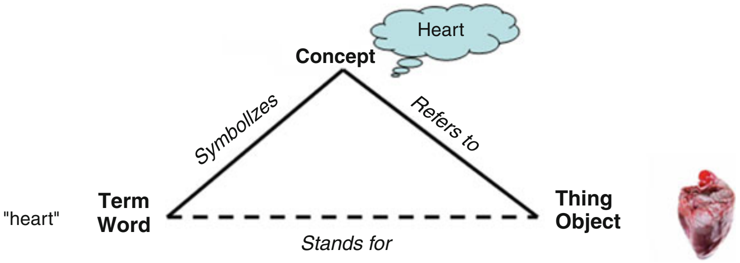


Fig. 2.2 Modified Ogden-Richards triangle

(Campbell et al. 1998): Thing or Object, Thought or Concept, Word (symbol) or Term (Figs. 2.1 and 2.2).

Within a person, the "heart" is a real thing or object. When thinking about this heart there is a thought in the brain referring to heart, which is referred to as a concept. This thought/concept is symbolised by the word/term/symbol "heart" in English.

2.4.2 Definition of Components According to (ISO 1087–1 2000) and (ISO 17115 2007)

2.4.2.1 Concept

A concept is "a unit of knowledge created by a unique combination of characteristics".

It may refer to a material thing (a car) or an immaterial entity (speed). It constitutes the apex of the Ogden-Richards triangle (Figs. 2.1 and 2.2).

It is symbolised by a designation.

2.4.2.2 Designation and Term

Designation is the representation of a concept by a sign, which denotes it. There are three types of designation: symbols, appellations and terms.

A term is the verbal designation of a general concept in a specific subject field, whereas an appellation is the verbal designation of an individual concept.

The term is the lower left point of the Ogden-Richards triangle and corresponds to the object (lower right point) expressed indirectly via the concept (Figs. 2.1 and 2.2).

2.4.2.3 Concept System

Intuitively, concepts can be placed in an organised system: for example, a closed fracture is a type of fracture.

If a terminological phrase is more complex than can be symbolised by a single term, such as “fracture”, it is necessary to define several concepts and their relationships.

A concept system is a set of concepts structured as a function of the relationships between them. This set of concepts and relationships is the basis of semantic representation.

There are two main types of relationship: hierarchical and associative.

A hierarchical relationship is a relationship between two concepts that may be either generic or partitive.

An associative relationship is a pragmatic relationship between two concepts having a non hierarchical thematic connection, by virtue of experience, as a causal, site or a temporal relationship. Most concept systems are based on generic relationships (symbolised by IS_A) or partitive relations (symbolised by PART_OF).

A generic relationship is a relationship between two concepts in which the intension (definition) of one concept includes that of the other concept plus at least one additional delimiting characteristic. For instance, the subordinate concept “Talus” has a generic relationship (IS_A) with the superordinate concept “Foot bone”. It is a foot bone, but has an additional characteristic (Fig. 2.3).

A partitive relationship is a relationship between two concepts in which one concept is the whole and the other is a part of the whole. For instance, the subordinate concept “Talus” has a partitive relationship (PART_OF) with the superordinate concept “Foot bone structure”. It is part of the bone structure of the foot (Fig. 2.3). In a generic relationship, the superordinate concepts named generic concepts have a narrower intension (definition) and lie at the top of the hierarchy and the subordinate concepts are specific concepts that are more precise and located at a lower level of the hierarchy.

In a partitive relationship, the superordinate concept known as the comprehensive concept is connected with co-ordinate concepts, which are at the same level of the hierarchy.

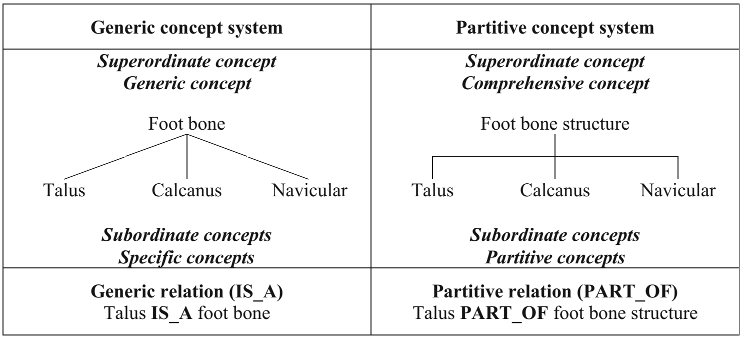


Fig. 2.3 Hierarchical concept systems

2.4.2.4 Definitions

A definition is the representation of a concept by a descriptive statement differentiating it from related concepts. The intensional definition of a concept is a definition describing the intension of a concept by stating the superordinate concept and its delimiting characteristics. (e.g. a femur diaphyseal fracture Is_A a fracture located on the femur diaphysis).

The extensional definition of a concept is the description obtained by grouping together all the subordinate concepts under a single criterion of subdivision. (e.g. noble gas : helium, neon, argon, krypton, xenon, radon).

2.4.3 Compositional Approaches for Concept Representation

Some simple concepts can be combined into a compositional concept representation. Let us take as an example "*Escherichia coli* pyelonephritis". It is possible to identify three categories, classes or axes of concepts: Topography with 'the pelvis or the kidney' (pyelonephr-), Morphology with 'infection' (-itis) and Etiology with '*Escherichia coli*'. The representation of this compound knowledge requires explicit description of the relationships between the components. In our example, "*Escherichia coli* pyelonephritis" can be represented as an infection (morphology) which "has_site" "the kidney" (topography) and which "has_cause" "*Escherichia coli*" (etiology).

The prevention of nonsense representations (such as liver fracture), requires the imposition of constraints between the relationships formalised as semantic links and the authorised components (formalised as categories, classes or axes of characterising concepts). For example, the semantic link "has_site" is authorised only between concepts characterising morphology and concepts characterising topography (Fig. 2.4).

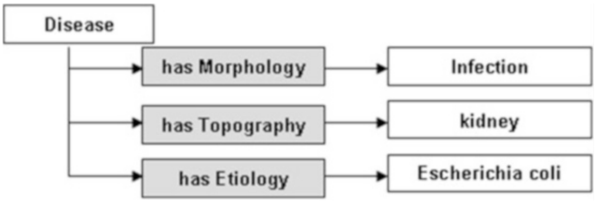


Fig. 2.4 Representation of the concept "*Escherichia coli* pyelonephritis"

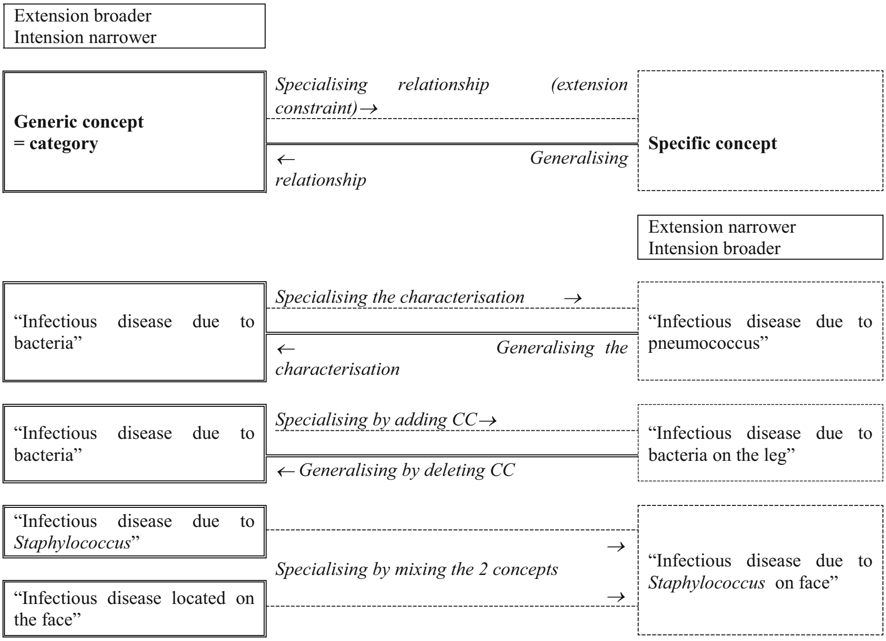

Extension broader
Intension narrower
Generic concept = category
Specialising relationship (extension constraint)→
← Generalising relationship
Specific concept
Extension narrower
Intension broader
"Infectious disease due to bacteria"
Specialising the characterisation →
← Generalising the characterisation
"Infectious disease due to pneumococcus"
"Infectious disease due to bacteria"
Specialising by adding CC→
← Generalising by deleting CC
"Infectious disease due to bacteria on the leg"
"Infectious disease due to Staphylococcus"
"Infectious disease located on the face"
Specialising by mixing the 2 concepts
"Infectious disease due to Staphylococcus on face"

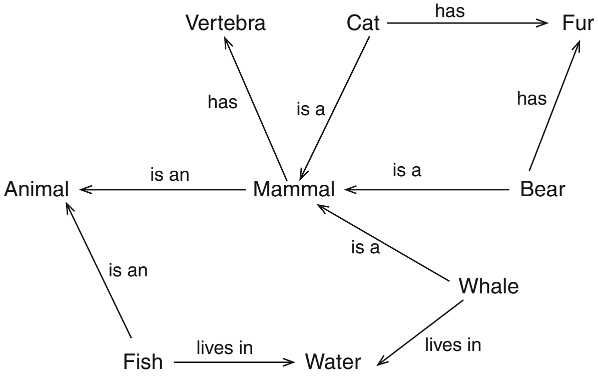
Fig. 2.5 Specialising/Generalising processes in a compositional system (*CC* composite characteristic)

The hierarchy, semantic links and compositional constraints, known as the categorial structure, of a set of elementary concepts defines the conceptual representation field. This field can be used to infer and to subsume automatically the subordinate concepts, as summarised in Fig. 2.5.

2.4.4 Formal Concept Representation

Various knowledge representation tools are available to support compositional approaches to concept representation. Knowledge representation tools are artificial intelligence tools for the representation of knowledge as symbols, facilitating

Fig. 2.6 Semantic network of C. Peirce



inference from knowledge elements to create new knowledge elements. The first formal representation proposed was called the Semantic or Frame network and was described in 1909 by C. Peirce (1909) as shown in Fig. 2.6.

It was developed into a graphical interface of first-order logic known as the Conceptual Graph by J Sowa (1984) and, more recently, into the Web Ontology Language, which combines the RDF/XML syntax format and description logic-based formal representation (Baader et al. 2005; Lacy 2005). An example is provided by the work of Schulz et al. (2011), using Bio Top upper level ontology:

> PathologicalEntity equivalent to PathologicalStructure or Pathological Disposition or Pathological Process

with

> All instances of PathologicalStructure are related to the anatomical objects where they occur via the relation PhysicalPartOf or by the more general relation physicallyLocatedIn.
>
> All instances of PathologicalDisposition are related to their bearers by the relation inheresIn.
>
> All instances of PathologicalProcess are related to the place where they occur by the relation hasLocus and to their participating entities by has Participant.

2.5 Terminological System Typology

A terminological system is a system organising the relationships between terms and concepts in a domain with, when appropriate, any associated rules, relationships, definitions and codes (EN ISO 1828 2012). The different types are named: terminology, nomenclature, thesaurus, vocabulary, classification, coding system, taxonomy and ontology.

A terminology is a set of designations belonging to a special language (ISO 1087–1 2000) related to the concepts of a specific domain (e.g. Terminologica Anatomica). Clinical and reference terminologies can be distinguished on the basis of their use.

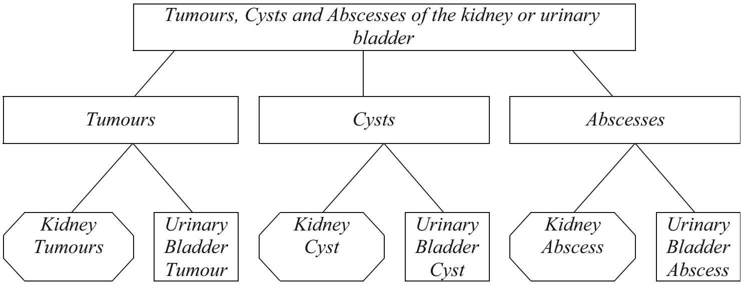


Fig. 2.7 Example of classification

A thesaurus is a dictionary of words in alphabetical order (with keywords and synonyms) organised to facilitate the retrieval and classification of documents, in an index, for example. In a vocabulary, terms are associated with definitions.

A nomenclature is an inventory of terms used to designate objects in a particular field, mostly when the system is based on user-specific rules rather than concepts.

A classification is an organisation of the exhaustive set of concepts of a domain, by necessary and sufficient conditions, such that each concept belongs to only one class. The classes are mutually exclusive and hierarchical (generic or partitive) and exhaustive, due to the creation of residual classes named “Not Elsewhere Classified” and “Not Otherwise Specified” (Fig. 2.7).

A taxonomy is a classification based exclusively on generic hierarchical relations.

A coding system is a combination of a set of concepts, a set of code values, and at least one coding scheme mapping code values to coded concepts (ISO 17115 2007). Codes are used by computers.

Codes may be meaningful, if a human can infer some knowledge from the code: for instance, the ICD 10 Congestive heart failure code I50.0 means that this disease can be found in chapter I, which relates to cardiovascular diseases. However, codes may be meaningless and entirely unlinked to any meaning (e.g. purely alphanumeric strings).

IDC 10 Typology

There are several types of healthcare terminological system. The International Classification of Disease (ICD 10) is a 21-chapter classification based on anatomy and aetiology concepts. Its codes are meaningful.

Volume 2 is an alphabetical index with synonyms, which may be considered to be a thesaurus. Its extension to oncology ICD-O makes it possible to create compositional representations with morphology. It is also a nomenclature. Finally, the coding rules in Volume 3 propose a semi-formal definition of the classes (De Keizer et al. 2000).

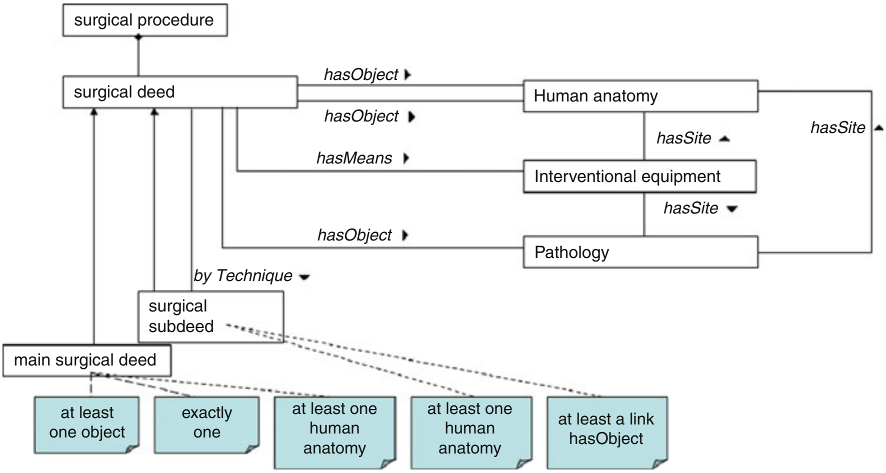


Fig. 2.8 Categorial structure for terminological systems of surgical procedures (Adapted from EN ISO 1828 2012)

Ontology is the study of what exists (the branch of metaphysics dealing with reality). Formal ontologies are theories that aim to provide precise mathematical formulations of the properties and relationships of certain entities.

Three levels of ontologies must be considered:

- Upper-level ontologies, representing the world: BFO, Bio Top, DOLCE, SUMO.
- Reference ontologies in a domain, such as FMA in anatomy, Galen for surgical procedures.
- Applied ontologies, such as SNOMED CT.

Compositional system representations and categorial structure (EN ISO 1828 2012) are semi-formal approaches to ontology. Categorial structures are minimal logic constraints for representing the concepts of a specific domain. For instance, for surgical procedures (Fig. 2.8), according to the categorial structure:

- The principal semantic categories are Anatomy, Deed, Device and Pathology, with two qualifier categories (cardinality and laterality).
- The semantic links are *by_technique*, *has_means*, *has_object*, *has_site* (with inverse):
  - *Has_object* is authorised between deed and anatomy or device or pathology,
  - *Has_site* is authorised between device or pathology and anatomy,
  - *Has_means* is authorised between deed and anatomy, device or pathology,
  - *By_technique* is authorised between deed and deed.
- The minimal constraints required are:
  - A deed and *has_object* shall be present,
  - Anatomy must always be present with either a *has_object* or with a *has_site*,

- The use of pathology is restricted to macroscopic lesions and to cases in which it can distinguish the procedure concerned from other procedures using the same deed and the same anatomy,
- When *by_technique* is used, the deed on the right side of the semantic link must conform to the previous rules.

The categorial structure makes it possible to ensure that new terms describing surgical procedures are associated with a formal definition consistent with a common template.

2.6 Desiderata for Terminological Systems

JJ. Cimino from the Columbian Presbyterian Medical Center in New York has defined 12 characteristics, known as desiderata, for terminological systems used in medical records (Cimino 1998).

1. The content must satisfy the user. To most users "What can be said" is more important than "how it can be said". Omissions are readily observed and timely, formal and explicit methods for plugging gaps are required.
2. The vocabulary must be concept-oriented. The unit of symbolic processing is the concept and each concept in the vocabulary should have a single, coherent meaning.
3. A concept's meaning cannot change and it cannot be deleted from the vocabulary, it is the concept permanence principle.
4. Concept identifiers must be meaningless. Concepts typically have unique identifiers (codes) and these should be non-hierarchical (see code-dependence), to allow for later relocation and multiple classification.
5. The system must be polyhierarchical, to allow multiple classification.
6. Concepts must have a semantic definition. For example, Streptococcal tonsillitis = Infection of the tonsil caused by *Streptococcus*.
7. The system must not have residual categories. Traditional classifications have rubrics that include NOS, NEC, Unspecified, Other, the meaning of which may change over time as new concepts are added to the vocabulary. These are not appropriate for recording data in an electronic health record.
8. The system must have multiple granularities. Different users require different levels of expressivity. A general practitioner might use myocardial infarction, whereas a surgeon may record acute anteroseptal myocardial infarction.
9. Although there may be multiple views of the hierarchy required to support different functional requirements and levels of detail, they must be consistent.
10. There is a crucial relationship between concepts within the vocabulary and the context in which they are used. Cimino defined three types of knowledge:

    - Definitional – how concepts define each another
    - Assertional – how concepts combine
    - Contextual – how concepts are used

11. Vocabularies must be designed to allow for evolution and change, to incorporate new advances in healthcare and to correct errors.
12. Where the same information can be expressed in different ways, a mechanism for recognising equivalence is required. This is redundancy recognition.

2.7 Terminologies in Action

Many terminologies have been designed, each for a specific purpose. They are used:

- To code patient data, in the context of health care, in epidemiological studies or public health;
- To index documents, including biomedical research articles;
- To represent entities in expert systems and decision support systems;
- To serve as an interface for data entry.

Several representations of a given condition may, therefore, co-exist. We will illustrate this phenomenon with the example of haemochromatosis. Haemochromatosis is a disorder that causes the body to absorb and to store too much iron. In the body, iron is incorporated into haemoglobin, which transports oxygen in the blood. Healthy people usually absorb about 10 % of the iron present in the food they eat. People with genetic haemochromatosis absorb about 20 % of the iron they ingest. The body has no natural way to rid itself of excess iron, so extra iron is stored in body tissues, especially the liver, heart and pancreas (source: www.niddk.nih.gov).

The accumulation of iron in body tissues may lead to:

- Osteo-articular symptoms, including joint pain and arthritis;
- Liver disease, including cirrhosis, cancer and liver failure;
- Heart disease, potentially leading to heart failure;
- Abnormal pigmentation of the skin;
- Damage to the pancreas, possibly causing diabetes;
- Impotence.

Symptoms tend to occur in men between the ages of 30 and 50, and in women over the age of 50. However, many people have no symptoms when they are diagnosed.

Genetic haemochromatosis is mostly associated with a defect in the HFE gene. HFE regulates the amount of iron absorbed from food. Two mutations, C282Y and H63D, are known to cause haemochromatosis. The genetic defect is present at birth, but symptoms rarely appear before adulthood.

Haemochromatosis may also be acquired, through blood transfusions, for example.

2.7.1 Terminologies and Their Use to Code Diseases

Many classification systems were designed in the seventeenth and eighteenth centuries. Nosologists tried to do for diseases what botanists had done for plants: to find the natural divisions between diseases, to discover the real essence of the diseases, and to embody this essence in a suitable definition. Thomas Sydenham (1624–1699) was one of the most famous nosologists. He said “It is necessary that all diseases be reduced to definite and certain species, and that, with the same care which we see exhibited by botanists in their phytology.”

The classification systems created during this period include Genera morborum (Linnaeus) and Nosologia Methodica (François Bossier de Lacroix also known as Sauvages). Nosology is the key to improving diagnosis and treatment. Sauvages saw nosology as a practical discipline providing practitioners with a compass to chart their voyages through the complex sea of symptoms.

From the nineteenth century onwards, increasing numbers of terminologies were created for practical purposes, such as the reporting of causes of death, particularly in England, for analyses of child mortality and the reporting of cases of plague in London. The need for accurate reporting of the causes of death (William Farr, Jacques Bertillon) led to the development of the International Classification of Diseases.

The International Classification of Diseases is a statistical classification dating back to the eighteenth century that is maintained by the World Health Classification. Its early revisions related exclusively to causes of death. Its scope was extended, in 1948, to include non-fatal diseases. Various versions of the International Classification of Diseases are now used in more than 50 countries, to code diagnoses in the DRG system or equivalent. The tenth revision was released in 1993. It contains 9876 items, allowing the coding of any case, thanks to the “not classified elsewhere” codes (e.g., “other disorders of mineral metabolism”) and the “not otherwise specified” codes (e.g. “disorders of mineral metabolism, unspecified”). Granularity varies between items, depending on statistical aspects. Moreover, classification criteria are included: for example E83.1 “Disorders of iron metabolism” is a subclass of E83 “Disorders of mineral metabolism”; it includes haemochromatosis but excludes iron deficiency anaemia. Haemochromatosis does not have its own class and cases of haemochromatosis are therefore simply coded as E83.1 “Disorders of iron metabolism” (Fig. 2.9).

Some diseases may be coded on the basis of their aetiology (B05.0: Measles complicated by encephalitis) or signs (G05.1: Encephalitis, myelitis and encephalomyelitis in viral diseases classified elsewhere).

The 11th version of the ICD will soon be released. Various models have been defined for the reporting of causes of death, the reporting of morbidity, disease coding in DRG systems, and primary care. Traditional medicine, including the various forms of Asian medicine, will also be represented in ICD 11. Moreover, ICD11 will be aligned with SNOMED CT.

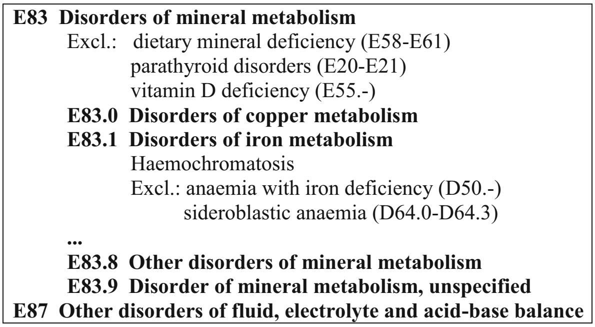


Fig. 2.9 Haemochromatosis in ICD 10

E83 Disorders of mineral metabolism
Excl.: dietary mineral deficiency (E58-E61)
parathyroid disorders (E20-E21)
vitamin D deficiency (E55.-)
E83.0 Disorders of copper metabolism
E83.1 Disorders of iron metabolism
Haemochromatosis
Excl.: anaemia with iron deficiency (D50.-)
sideroblastic anaemia (D64.0-D64.3)
...
E83.8 Other disorders of mineral metabolism
E83.9 Disorder of mineral metabolism, unspecified
E87 Other disorders of fluid, electrolyte and acid-base balance

SNOMED CT (http://www.ihtsdo.org/snomed-ct/) is a clinical reference terminology. It is a comprehensive terminological system for coding clinical information (283,000 concepts, 732,000 terms and 923,000 relationships).

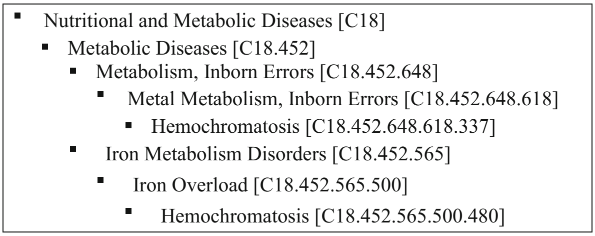
2.7.2 Terminologies for Indexing

There is a need for controlled vocabularies suitable for use in the indexing and cataloguing of biomedical publications. These terminologies must be thesauri containing links showing the relationships between related terms and providing a hierarchical structure facilitating searching at various levels of specificity, from "narrower" to "broader" terms. They correspond to a more or less limited list of terms encompassing synonyms, to facilitate information retrieval.

Medical Subject Headings (MeSH) is the controlled-vocabulary thesaurus used for indexing articles for PubMed. MeSH was designed in 1960 by the US National Library of Medicine. The 2012 edition of MeSH contains 26,581 descriptors (subject headings), which are used to index documents in an unambiguous manner. The descriptors are organised into several modules covering all the domains of biomedicine: Anatomy, Organisms, Diseases, Chemicals and Drugs, Analytical, Diagnostic and Therapeutic Techniques and Equipment, Psychiatry and Psychology, Phenomena and Processes, Disciplines and Occupations, Technology, Industry, Agriculture, Anthropology, Education, Sociology and Social Phenomena, Humanities, Information Science, Named Groups, Health Care, Publication Characteristics, Geographical. MeSH is a directed acyclic graph, in which a term may have more than one parent, for example

- Urinary lithiasis *is a* Lithiasis,
- Urinary lithiasis *is a* Urologic disease.

In MeSH, "Haemochromatosis" is categorised as "Metal Metabolism, inborn errors". In fact, we know that the relationship between "Haemochromatosis" and "Metal Metabolism, inborn errors" should actually be "is generally a" rather than "is a", because haemochromatosis is acquired in some cases, through multiple



- Nutritional and Metabolic Diseases [C18]
  - Metabolic Diseases [C18.452]
    - Metabolism, Inborn Errors [C18.452.648]
      - Metal Metabolism, Inborn Errors [C18.452.648.618]
        - Hemochromatosis [C18.452.648.618.337]
    - Iron Metabolism Disorders [C18.452.565]
      - Iron Overload [C18.452.565.500]
        - Hemochromatosis [C18.452.565.500.480]

Fig. 2.10 Haemochromatosis in MeSH

blood transfusions. Such relationships, although not taxonomic, are of interest for information retrieval purposes. For this reason, “Haemochromatosis” is also categorised as “Iron Overload” (Fig 2.10).

2.7.3 Terminologies in Decision Support Systems

Reasoning in decision support systems may be based on taxonomies, in addition to rule-based inferences. Taxonomies support specialisation and generalisation (from the more general to the more specific and vice versa).

Quick Medical Reference (QMR) is a decision support system for assisted diagnosis in medicine, not restricted to a specific domain. It uses a terminology to represent diseases, signs and symptoms. The QMR terminology is organized as a taxonomy, in which haemochromatosis is a subclass of cirrhosis. In this representation, the system focuses on the liver lesions caused by iron overload. Cirrhosis is a complication of haemochromatosis. The relationship between haemochromatosis and cirrhosis is therefore not taxonomic and actually means “may be found when”.

2.7.4 Terminologies for Data Entry

Interface terminologies are used to facilitate data entry into electronic medical records. They link user’s descriptions to structured data elements in a reference terminology (Rosenbloom et al. 2006).

During the development of a domain-specific interface terminology based on a reference terminology, the domain concepts are identified and mapped to the reference terminology concepts. This creates a subset of the reference terminology (Bakhshi-Raiez et al. 2010). This subset covers the needs of the user and is not directly displayed to users, instead being presented in terms or descriptions familiar to the user.

SNOMED CT may be used as an interface terminology for data entry as it includes many terms for each concept and subsets for data entry are made available.

The VCM iconic language (Visualization of Concepts in Medicine) (http://vcm.univ-paris13.fr/svcm) proposes icons for the graphic representation of the main physiopathological states: patient characteristics (e.g. age, sex), symptoms, diseases, antecedents, risks, various classes of treatment, medical follow-up procedures, health professionals and medical speciality, and medical knowledge. These icons are created by combining several elements from a lexicon of shapes, colours and pictograms, according to simple grammar rules. A health professional can usually learn the VCM lexicon and grammar in a few hours. The VCM language makes the *is-a* relations present in other terminologies visually explicit. For example, a coronary disease is a cardiac disease, but this is not explicit in a visual search because the "coronary disease" term does not include the word "heart" or the prefix "card-". By contrast, the VCM icon for coronary disease makes the relationship with cardiac disease explicit because it includes the heart pictogram. This iconic language is accompanied by a graphical silhouette (Mister VCM) making it possible to organise, in a restricted space, a set of VCM icons corresponding to coded data by anatomy and aetiology.

2.7.5 Conclusions

Each terminology is useful in a given context, but not universally valid. The organisation of the concepts reflects the intended objective. It is useful to find out the aim of a terminology before using it, particularly when hierarchical structures are used in algorithmic process. The underlying question is: Does this resource describe the domain and the concepts needed to achieve the desired goal adequately.

2.8 Aligning Terminologies

2.8.1 Alignment Methods

As detailed above, in health there are almost as many different terminologies, controlled vocabularies, thesauri and classification systems as there are fields of application (Shvaiko and Euzenat 2013).

Given the enormous number of terminologies, existing tools, such as search engines, coding systems and decision support systems, have a limited capacity for dealing with "syntactic" and "semantic" divergences, despite their large storage capacities and ability to process data rapidly. Faced with this reality and the increasing need to allow co-operation with/between the various health actors and their related health information systems, there seems to be a need to link and connect these terminologies, to make them "interoperable". Alignment techniques are of particular importance because the manual creation of correspondences between concepts or between terms is extremely time-consuming. There are two major dimensions for similarity: the syntactic dimension and the semantic dimension. The syntactic dimension is based on lexical methods and the semantic dimension is based on structural and semantic properties of terminologies.

2.8.1.1 Lexical Methods

Lexical methods are based on the lexical properties of terms. These methods are straightforward and constitute a trivial approach to identifying correspondences between terms. The use of such methods to achieve mapping in the medical domain was driven by the similarly of the terms included in many terminologies.

String-Based Methods

In these methods, terms (or labels) are considered as sequences of characters (strings). A string distance is calculated to determine the degree of similarity between two strings. In some of these methods, the order of the characters is not important. Examples of such distances, also used in the context of information retrieval, are: the Hamming distance, the Jaccard distance, the Dice distance. Another family of appropriate measures, known as the "Edit distance" takes into account the order of characters. Intuitively, an edit distance between two strings is defined as the minimum number of character insertions, deletions and changes required to convert one string into another. The Levenshtein distance is one example of such a distance. It is the edit distance with all costs (character insertions, deletions and changes) equal to 1. This measure is also frequently used for spelling errors. For example: asthmma vs. asthma (insertion of one character), astma vs. asthma (deletion of one character) and ashtma vs. asthma (inversion of two characters). However, these methods can only quantify the similarity between terms or labels. They therefore provide low estimates of similarity between synonymous terms with different structures. For example, the words "pain" and "ache" are synonyms. They are thus semantically related and mean the same thing, but none of the distances presented above can identify any links between these two terms. Conversely, these methods find significant similarity between terms that are actually different (false positive), such as "Vitamin A" and "Vitamin B".

Language-Based Methods

In these methods, terms are considered as words in a particular language. NLP tools are used to facilitate the extraction of meaningful terms from a text. These tools exploit the morphological properties of words. Methods based on normalisation processes can be distinguished from those making use of external knowledge resources, such as dictionaries.

Normalisation Methods

Each word is normalised to a standardised form that is easy to recognise. Several linguistic software tools have been developed for the rapid retrieval of a normal form of strings: (i) tokenisation involves segmenting strings into sequences of tokens, by eliminating punctuation, cases and blank characters; (ii) the stemming process involves analysing the tokens derived from the tokenisation process, to reduce them to a canonical form; (iii) stop words elimination involves removing all the frequent short words that do not affect the sentences or the labels of terms, phrases such as "a", "Nos", "of"...etc.

External Knowledge-Based Methods

These methods use external resources, such as dictionaries and lexicons. Several linguistic resources have been developed to identify possible mappings between terminologies. These methods form the basis of the lexical tools used by the UMLSKS API (https://uts.nlm.nih.gov/home.html). They were combined with synonyms from other external resources to optimise mapping to the UMLS. Another external resource largely used in the biomedical field is the lexical database WordNet (http://wordnet.princeton.edu/).

Examples

Exact Match The Orphanet disease "Glycogen storage disease type 2" has an exact match with the SNOMED notion "Glycogen storage disease, type II": "Glycogen storage disease type 2" is a synonym of the MeSH descriptor "Glycogen storage disease, type II", which is itself an exact match for the SNOMED notion "Glycogen storage disease, type II".

Alignment by Combination The Orphanet term "Diabetic embryopathy" is aligned with two MeSH descriptors "Diabetes mellitus" and "Fetal diseases": with NLP tools "Diabetic" is matched with the MeSH descriptor "Diabetes mellitus"; and "Embryopathies" is a MeSH synonym of "Fetal diseases" (which is, here, an exact match).

2.8.1.2 Structural Methods

These methods use the structural properties of each terminology to identify all the possible correspondences between terms. They consider terminologies as graphs in which the nodes represent terms and the edges represent relationships between these terms established in the terminology. Most medical terminologies can be represented as graphs. Furthermore, these techniques can also be combined with lexical techniques. Together with the structural properties of each terminology,

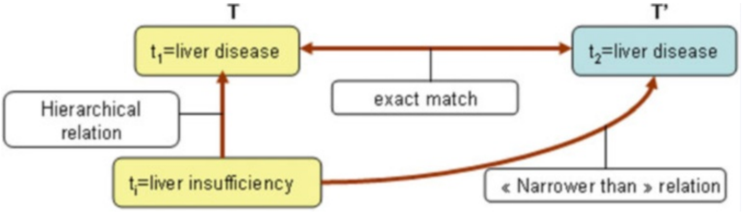


Fig. 2.11 NT alignment

semantic methods also use semantic similarities to find the closest term. The principal technique involves calculating the number of edges between terms, to determine a distance between them. The best known distance estimating similarity is the Wu-Palmer distance, which is defined according to the distance between two terms in the hierarchy and their positions with respect to the root. Unlike these traditional edge-counting approaches, other methods, such as Lin similarity methods, estimate similarity from the maximum amount of information shared by two terms in a hierarchical structure. These similarities can be used to find possible connections between terms or concepts from different hierarchical terminologies, such as MeSH or SNOMED INT, for example. This approach is based on hierarchical relationships and is used to align the remaining terms not mapped by the lexical approach. This mapping provides two types of correspondences:

- Narrow-Mapping: when the remaining term has at least one child (hierarchical relationship narrower than) mapped to at least one term.
- Broad-Mapping: when the remaining term has at least one parent (hierarchical relationship broader than) mapped to at least one term.

"Narrower Than" (NT) Relationships

If a term t_1 from the terminology T has an exact lexical match with a term t_2 from the terminology T', then each term t_i of T narrower than t_1 is narrower than t_2 in T'. A NT relation links t_i in T to t_2 in T' (Fig. 2.11).

"Broader Than" (BT) Relationship

If a term t_1 from the terminology T has an exact lexical match with a term t_2 from the terminology T', then each term t_i of T broader than t_1 is broader than t_2 in T'. A BT relationship links t_i in T to t_2 in T'.

However, an evaluation, generally a manual evaluation, is required to ensure that the alignment is of high quality.

2.8.2 *The UMLS*

In 1986, US National Library of Medicine (NLM) launched the Unified Medical Language System (UMLS) project. According to Donald Lindberg, Director of the NLM, the objective was to build a vocabulary, a language linking the biomedical literature with observations on the patient, and educational applications in the school, a language connecting these areas.

The UMLS project addresses semantic heterogeneity issues in the biomedical domain. It provides terminological resources that have been made available to the community (Knowledge Sources). Since 1990, the UMLS project has produced annual editions of tangible products that are now regularly used by their intended audience. Assessments of the value of the UMLS products by more disinterested observers are required, but an increasing array of operational systems are making use of one or more of the UMLS Knowledge Sources or lexical programs (Lindberg et al. 1993). The UMLS resources include:

- The Metathesaurus (1990), including a large set of terminologies used biomedical domain and describing relationships between the terms;
- The Semantic Network (1990) is a set of semantic types representing the broad categories of the domain. They are used to categorise the Metathesaurus concepts;
- The Specialist Lexicon (1994) provides the lexical information required for natural language processing. It includes commonly occurring English words and biomedical vocabulary. The Lexicon entry for each word or term records the syntactic, morphological and orthographic information used with associated NLP tools.

2.8.2.1 The UMLS Metathesaurus

The Unified Medical Language System® (UMLS®) Metathesaurus® is a large, multi-purpose, multilingual thesaurus that contains millions of biomedical and health-related concepts, their synonymous names and their relationships. The Metathesaurus includes over 150 electronic versions of classifications, code sets, thesauri, and lists of controlled terms in the biomedical domain. These are the source vocabularies of the Metathesaurus. (http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html)

The Metathesaurus is organized by concept, or meaning. It links alternative names and views of the same concept from different source vocabularies and identifies useful relationships between different concepts. As such, the UMLS Metathesaurus transcends the specific thesauri, codes and classifications it encompasses.

The UMLS Metathesaurus includes most of the terminologies used in medicine, such as the International Classification of Diseases and SNOMED. It includes different versions of the terminologies, such as ICD9, ICD10, ICD9-CM, and different languages (ICD in French, Spanish, German, Russian, etc.). Each concept has a unique identifier, known as the Concept Unique Identifier (CUI). All the relationships existing between two terms in the source terminologies are represented. All information, terms, concepts and relationships are presented in a unified format.

We can illustrate this with craniostenosis.

– Many source terminologies contain this concept: ICD-10, ICPC, MedDRA, MeSH, OMIM, Read Codes, SNOMED CT
– A definition can be found in MeSH: Premature closure of one or more sutures of the skull;
– Synonymous terms are clustered into a single CUI C0010278 with a preferred label "Craniosynostosis" (Preferred Term). The synonymous terms include Craniostenosis (ICD, ICPC, OMIM, SNOMED CT), Craniosynostosis syndrome (SNOMED CT), Synostosis (cranial) (CRISP), Premature closure of cranial sutures (MedDRA, SNCT), Congenital ossification of cranial sutures, Congenital ossification of sutures of skull, Premature cranial suture closure (SNOMED). We also have abbreviations, such as CRS, CSO and CRS1 (OMIM). In addition, the UMLS provides translations into several languages (e.g. Spanish, German, French).
– The Metathesaurus provides a list of related terms, either more specific, such as Craniosynostosis, type 1 (OMIM), or syndromes such as Hurst syndrome (C0014077) Christian syndrome 1 (C0795794) or SCARF (Skeletal abnormalities, cutis laxa, craniostenosis, ambiguous genitalia, psychomotor retardation, facial abnormalities) syndrome (C0796146);
– The relationships between the concept "Craniostenosis" and other concepts in source terminologies are retained.

The UMLS Metathesaurus contains more than 9,000,000 terms, 2,000,000 concepts and 22,000,000 relationships between concepts. The data correspond to 152 terminologies and 19 different languages.

2.8.2.2 The UMLS Semantic Network

The UMLS Semantic Network consists of (i) about 130 broad categories, or semantic types, providing a consistent categorisation of all concepts represented in the UMLS Metathesaurus®, and (ii) a set of semantic relationships between semantic types (network). The semantic types are of two kinds:

– "Entity" encompasses physical objects (e.g. plants, animals, anatomical structures, chemicals) and conceptual entities (e.g. spatial entities, temporal entities, signs and symptoms);
– "Event" includes activities (e.g. therapeutic procedures, behaviour), phenomena and processes (e.g., biological functions, diseases).

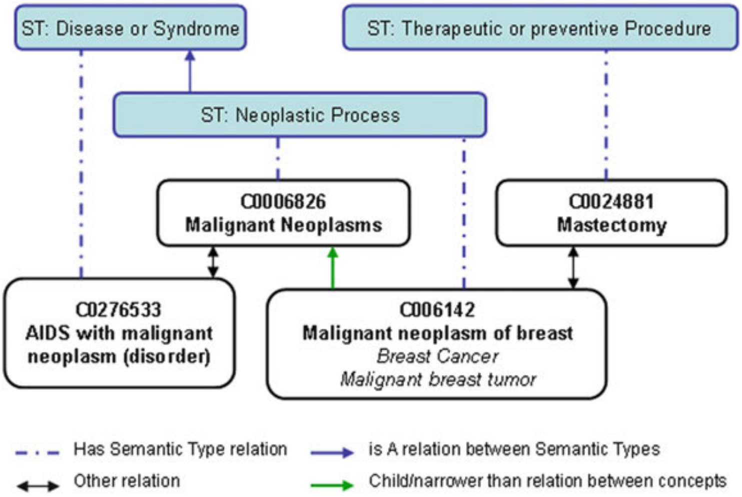

ST: Disease or Syndrome
ST: Therapeutic or preventive Procedure
ST: Neoplastic Process
C0006826
Malignant Neoplasms
C0024881
Mastectomy
C0276533
AIDS with malignant neoplasm (disorder)
C006142
Malignant neoplasm of breast
Breast Cancer
Malignant breast tumor
Has Semantic Type relation
Other relation
is A relation between Semantic Types
Child/narrower than relation between concepts

Fig. 2.12 Concept categorisation in the UMLS

Each semantic type is given a definition.

Example of Semantic Types

UI: T190

STY: Anatomical abnormality

ABR: anab

STN: A1.2.2

DEF: An abnormal structure, or one that is abnormal in size or location.

UN: Use this type if the abnormality in question can be either acquired or congenital abnormality. Neoplasms are not included here. These are given the type ‘Neoplastic Process’. If an anatomical abnormality has a pathological manifestation, then it will additionally be given the type ‘Disease or Syndrome’, e.g., “Diabetic Cataract” will be double-typed for this reason.

HL: {isa} Anatomical Structure; {inverse_isa} Congenital Abnormality.

The Semantic Network represents 54 relationships. The links between the semantic types provide the structure for the network. The primary link between the semantic types is the “is_a” link, which establishes the hierarchy of types within the Semantic Network. There are also non-hierarchical relationships (e.g., treats, diagnoses) that represent useful and important relationships in the biomedical domain (e.g., drugs treat diseases).

The Semantic Network provides a broad categorisation of Metathesaurus concepts. At least one semantic type is assigned to each concept in the Metathesaurus. For example, mastectomy is categorised as a therapeutic or preventive procedure (Fig. 2.12).

2.8.2.3 SPECIALIST Lexicon

The SPECIALIST Lexicon was developed for the SPECIALIST Natural Language Processing (NLP) System. It consists of a set of lexical entries, with one entry for each spelling or set of spelling variants in a particular part of speech. Lexical items may be multi-word terms if the term is determined to be a lexical item by its presence as a term in general English or medical dictionaries, or in medical thesauri, such as MeSH®. The Lexicon also includes acronyms and abbreviations.

A set of 20,000 words form the core words entered. This core is derived from the UMLS Test Collection of MEDLINE® abstracts, together with words that appear in both the UMLS Metathesaurus and Dorland's Illustrated Medical Dictionary. It also includes words from the general English vocabulary and the 10,000 most frequently used words listed in The American Heritage Word Frequency Book and the list of 2,000 words used in definitions in Longman's Dictionary of Contemporary English.

(http://www.nlm.nih.gov/pubs/factsheets/umlslex.html).

The lexicon entry for each word or term records the syntactic, morphological, and orthographic information needed for NLP. For example, the verb "to treat" is associated with "treats", the third person singular form in the present tense, "treated" the past and past participle form, and "treating" the present participle form (McCray et al. 1994).

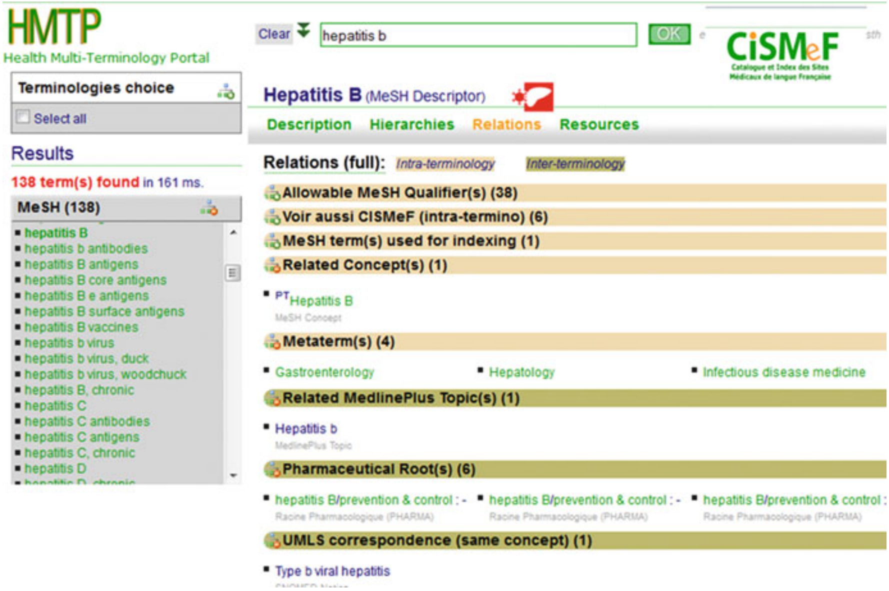
2.8.3 The Health Multiple-Terminologies Portal

Several terminological resources are aligned in the Metathesaurus® database of the UMLS. These relationships, via CUIs, may be exploited for semantic interoperability between medical applications. However, some terminologies are not included in the UMLS.

The Health Multi-Terminology Portal (HMTP) was created by the CISMeF team at Rouen University Hospital. It allows users to navigate through several terminologies (Fig. 2.13). The MeSH is available online, without login. After free registration on the website, users have access to the hierarchies of the following freely available terminologies and ontologies: ATC, CCAM, CIF, ICD-10, CISP-2, Cladimed, DRC, FMA, IUPAC, LOINC, LPP, MedlinePlus, NCCMERP, PSIP Taxonomy, SNOMED Int., VCM, WHO-ART and WHO-ICPS. The HMTP includes 32 terminologies (11 from the UMLS and 21 non-UMLS), for some of which, such as MedDRA and Orphanet, a licence fee must be paid.

When the user enters a term via the interface (Fig. 2.13), he or she can visualise, for this term:

- Its description ("Description" link);
- The list of the terminologies in which the term exists;
- The different hierarchies in which the term exists ("Hierarchies" link), with a visualisation of the complete trees or reduced trees with only the direct subsumers and subsumees;



HMTP
Health Multi-Terminology Portal

Clear hepatitis b OK

CiSMeF
Catalogue et Index des Sites Médicaux de langue Française

Terminologies choice
□ Select all

Results
138 term(s) found in 161 ms.

MeSH (138)
- hepatitis B
- hepatitis b antibodies
- hepatitis B antigens
- hepatitis B core antigens
- hepatitis B e antigens
- hepatitis B surface antigens
- hepatitis B vaccines
- hepatitis b virus
- hepatitis b virus, duck
- hepatitis b virus, woodchuck
- hepatitis B, chronic
- hepatitis C
- hepatitis C antibodies
- hepatitis C antigens
- hepatitis C, chronic
- hepatitis D

Hepatitis B (MeSH Descriptor)

Description Hierarchies Relations Resources

Relations (full): Intra-terminology Inter-terminology

Allowable MeSH Qualifier(s) (38)
Voir aussi CISMeF (intra-termino) (6)
MeSH term(s) used for indexing (1)
Related Concept(s) (1)
- PT Hepatitis B
  MeSH Concept
Metaterm(s) (4)
- Gastroenterology
- Hepatology
- Infectious disease medicine
Related MedlinePlus Topic(s) (1)
- Hepatitis b
  MedlinePlus Topic
Pharmaceutical Root(s) (6)
- hepatitis B/prevention & control : -
  Racine Pharmacologique (PHARMA)
- hepatitis B/prevention & control : -
  Racine Pharmacologique (PHARMA)
- hepatitis B/prevention & control :
  Racine Pharmacologique (PHARMA)
UMLS correspondence (same concept) (1)
- Type b viral hepatitis

Fig. 2.13 Interterminology and intraterminology relationships for “*hepatitis B*”

– Its intraterminology relationships with other terms in the chosen terminology (“Relations” link);
– Its interterminology relationships deduced with an UMLS Metathesaurus® alignment (“Relations” link);
– Its interterminology relationships deduced with an exact lexical match of the term (“Relations” link).

The HMTP has a multilingual version, the European Health Terminology/Ontology Portal (http://www.ehtop.eu/). EHTOP includes the 32 terminologies of the HMTP and the same search and navigation functionalities. Only the ICD10 is freely accessible in 11 languages.

2.9 Conclusion

The organisation of diseases into classifications initially facilitated analyses of the causes of death and the first calculations of mortality statistics. With the informatisation of health data, the use of standardised terminologies for the recording of medical information is essential, because it enables the reuse of this information for various goals (e.g. decision support, information retrieval).

Terminological systems have evolved. The first systems organised the terms into lists with or without hierarchical relationships. They were followed by systems allowing the composition of concepts, with a gain in expressivity. The most recent systems use formal definitions of concepts to enable the computer to reason on the basis of these concept definitions and to improve semantic representation. The SNOMED CT ontology is a terminological system increasingly widely adopted by developers of health information systems and is one of the cited terminological references for the implementation of interoperability frameworks (HL7).

2.10 For More Information

This chapter contains many links relating to this topic. We recommend that readers systematically consult the websites mentioned in this chapter.

Many terminological systems exist and only a few have been used as examples in this chapter. Most are accessible via the Internet. We provide a non-exhaustive list of Internet links to access them.

The classification of the world health organisation (WHO http://who.int/classifications/en/)

where you will find:

- International classification of disease, 10th revision (ICD10)
- International classification of disease, 11th revision (ICD11)
- International classification of disease for oncology (ICD-O)
- International classification of primary care, 2nd edition (ICPC)
- International classification of functioning, disability and health (ICF)
- International classification of health interventions (ICHI)
- Anatomical therapeutic classification (ATC) for drugs at http://www.whocc.no/atc_ddd_index/
- WHO Adverse Reaction Terminology (WHO-ART) at http://www.umc-products.com/
- International Classification for Patient Safety (ICPS) at http://www.who.int/patientsafety/en

The reference terminologies

For medicine: SNOMED CT: http://www.ihtsdo.org/

For radiology:

RADLEX: http://www.rsna.org/radlex/

SNOMED Dicom Microglossary: http://www.all-acronyms.com/cat/7/SDM/SNOMED_DICOM_Microglossary/964404

For anatomical pathology

ADICAP: http://www.adicap.asso.fr/

For psychiatry

Diagnostic and Statistical Manual of Mental Disorders IV (DSM): http://www.psych.org/mainmenu/research/dsmiv.aspx

For biological laboratory tests

Logical Observation Identifiers Names and Codes (LOINC): http://loinc.org/

For procedures

Classification Commune des Actes Medicaux (CCAM): http://www.ameli.fr/accueil-de-la-ccam/index.php

Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS4): http://www.connectingforhealth.nhs.uk/systemsandservices/data/clinicalcoding/codingstandards/opcs4

Procedure Coding System (PCS): www.cms.hhs.gov/ICD9ProviderDiagnosticCodes/08_ICD10.asp

For adverse drug reactions

Medical Dictionary for Regulatory Activities (MedDRA): http://www.ich.org/products/meddra.html

For anatomy

Foundational Model of Anatomy (FMA): http://sig.biostr.washington.edu/projects/fm/AboutFM.html

For genetics

Gene Ontology (GO): http://www.geneontology.org/

Human Genome Organisation (HUGO): http://bioportal.bioontology.org/ontologies/45082

The ontology portals

Open Biological and Biomedical Foundation (OBO Foundry): www.obofoundry.org

Bioportal: bioportal.bioontology.org

Exercise

Q1 Search for Haemochromatosis in SNOMED CT, using the following browser: http://vtsl.vetmed.vt.edu/. By looking at the definition of this concept and the set of its antecedents, can you say what haematochromatosis is?

Q2 Do you agree with the SNOMED CT representation of haematochromatosis?

R1 Enter “Haemochromatosis” in the description tab and select the concept “Haemochromatosis (disorder)”. Haemochromatosis “has for causal agent” iron.
It is an iron overload, a disorder of iron metabolism, a disorder related to the excess of a trace element. It is a disorder of mineral metabolism, a disorder related to excess intake of micronutrients, a disorder of hyperalimentation, a nutritional disorder.

R2 We can consider whether “disorder of hyperalimentation” is a definition that matches haemochromatosis correctly. This definition was certainly automatically generated (automatic treatment based on concept definitions), hence the need to develop methods for auditing large terminological systems to ensure medical consistency.

References

Baader F, Horrocks I, Sattler U (2005) Description logics as ontology languages for the semantic web. In: Stephan W, Hutter D (eds) Mechanizing mathematical reasoning. Springer, New York

Bakhshi-Raiez F, Ahmadian L, Cornet R et al (2010) Construction of an interface terminology on SNOMED CT. Generic approach and its application in intensive care. Methods Inf Med 49 (4):349–359

Campbell KE, Oliver DE et al (1998) Representing thoughts, words, and things in the UMLS. J Am Med Inform Assoc 5:421–443

Cimino JJ (1998) Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 37(4–5):394–403

de Keizer NF, Abu-Hanna A, Zwetsloot-Schonk JH (2000) Understanding terminological systems I; terminology and typology. Methods Inf Med 39(1):16–21

EN ISO 1828 (2012) Health informatics – categorial structure for classifications and coding systems of surgical procedures

ISO 1087–1 (2000) Terminology work – vocabulary – part 1: theory and application

ISO 17115 (2007) Health informatics – vocabulary for terminological systems

Lacy LW (2005) Chapter 10. OWL: representing information using the web ontology language. Trafford Publishing, Victoria

Lindberg DA, Humphreys BL, McCray AT (1993) The unified medical language system. Methods Inf Med 32(4):281–291

McCray AT, Srinivasan S, Browne AC (1994) Lexical methods for managing variation. In: Biomedical terminologies, proceedings of the 18th annual symposium on computer applications in medical care, pp 235–239

Ogden CK, Richards IA (1923) The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism. K. Paul, Trench, Trubner/Harcourt, Brace, London/New York

Peirce CS (1909) Existential graphs, MS 514. Eprint of existential graphs MS 514 with commentary by John F. Sowa

Rosenbloom ST, Miller RA et al (2006) Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc 13(3):277–288

Schulz S, Spackman K et al (2011) Scalable representations of diseases in biomedical ontologies. J Biomed Semantics 2(suppl 2):S6

Shvaiko P, Euzenat J (2013) Ontology matching: state of the art and future challenges. EEE Trans Knowl Data Eng 25(1):158–176

Sowa JF (1984) Conceptual structures: information processing in mind and machine. Addison-Wesley, Reading