

A conceptual model for a terminological system for healthcare in Sweden

Allan Junfors^{1,3}, Sven-Bertil Wallin^{1,4}, Anders Thurin^{2,5}, Mats Carlsson²

¹ Swedish Institute for Health Services Development,(SPRI),Stockholm, Sweden

² Dept of Medical Informatics, Linköping, Sweden

³ Dept of Computer and System Sciences/KTH, University of Stockholm, Sweden

⁴ AU-Gruppen AB, Stockholm, Sweden

⁵ Dept of Clinical Physiology, Vrinnevi Hospital, Norrköping, Sweden

*Correspondence to: Allan Junfors,Swedish Institute for Health Services Development,
PO Box 70487, S-107 26 Stockholm, Sweden, Tel: +46 8 702 46 29, Fax: +46 8 702 47 99

Abstract. This report describes goals, requirements and information structure of a planned terminological system for the healthcare sector of Sweden. A conceptual model developed in the Spriterm-project in Sweden is presented and its properties is discussed in relation to published criteria for a good terminology and also in comparison with other similar models.

1. Background

Information Technology makes it possible to handle a vast amount of information in different kinds of healthcare information systems. Medical information can be transferred by messages between information systems, both within and between organisational units, nationally and internationally.

One crucial problem in this context is that information exchange between computers must be unambiguous and strictly specified. One way to achieve this is through a common repository of all relevant concepts concerning history, investigations, diseases, therapy, planning etc., possible to use in medical records.

Computers will also be able both to interpret medical texts and extract medical concepts from their expression in natural language, and to generate natural language from internal representations. Both these functions can be supported by a terminological system. To deliver these functions to Swedish healthcare a national terminological system consisting of clinically useful terms in Swedish, mapped to a controlled conceptual representation is required.

For international communication the system should also be internationally harmonised and include possibilities to use terms in different languages (a multilingual terminological system). To avoid the ambiguity of natural medical language, an artificial coding system is to be used when a precise identification of message content is needed, and also such codes should be mapped to the conceptual representation.

Information structure and content, functionality, technical design must be done in harmony with healthcare professionals and with constant consideration of end-user needs and clinical usefulness.

In this context a conceptual model has been developed to form the basis for a concept system according to goals and requirements here mentioned.

2. Terminological system

A terminological system is designed to fulfil a defined goal for use in special professional areas which delimit the subject fields, and define what is relevant regarding information content (target concepts, target categories and terms), and logical structure (semantic relations and their associated domain). The system also has to be designed in a technical structure and implemented in a technical environment. A terminological system is definitely more than just a list of terms.

The physical end-product (the terminological system) can be distributed as a special application including a terminological database with functions for retrieval of concepts and terms,

presentation and other types of processing. This kind of an application has been called a "terminology server"[4]

Building a terminological system around a conceptual representation will be essential to meet end-user requirements.

2.1. Purpose of a terminological system in healthcare

A terminological system should in our opinion be designed to support:

Primarily:

- unambiguous use of terms in computerized medical records systems within different clinical environments in helthcare
- sharing of information between different health care professionals in the care of a patient
- transfer of information

Secondarily:

- collection and aggregation of data for quality revision, clinical research, follow-up, analysis of results
- collection of epidemiological and other information for local, regional, national and international comparisons
- individual care-planning for a patient
- planning of the health care organization
- cost calculations and billing
- extraction of information e.g. to build knowledge bases
- knowledge based / decision support systems
- tools for development of application-systems such as data dictionaries

2.2. Requirements on a terminological system

To support these requirements, ideally, a computerized terminological system for healthcare should be an unambiguous, controlled terminology authorized and supported by the profession, and accepted and used in an uniform way by every professional working in the field. The terminological system is also required to make it easier to communicate information between organisational units locally, regionally, nationally and internationally. This is a very delicate, complex and costly task to reach if ever possible, but it outlines the direction to work towards.

Much work is needed by healthcare professionals to select terms, and to get a structure for authorization and support by the profession, as well as national and international harmonization.

The information needed normally already exists in other sources as classifications, nomenclatures, but they are developed for other purposes and in other structures. This means that the information has to be revised and converted according to the needs and the structure of the terminological system to avoid redundancy, ambiguity and misclassification. You cannot just integrate the information in your terminological system in an uncritical way. Relations to terms and codes in other sources may instead be kept by mapping or transformation.

The best way to reach this, as we see it is to build the system around semantic relations between concepts in the concept system. Thereby it is possible to implement more knowledge about concepts than just generic hierarchies - each concept can be defined by its relations to other concepts in the system instead of by complicated verbal descriptions. Verbal definitions also suffer from the difficulties to reach consensus about wording. Instead, every concept can be inserted "in its right place on the map".

Cimino et al. [2], consider their MED-system a hybrid between terminology and knowledge. Our opinion is that a well-balanced mix of terms, concepts and knowledge will give the best foundation for clinical usefulness.

Another question is how much knowledge in the form of semantic links is desirable? What is the borderline between a knowledge-system and a terminological system?

Cimino et al.[2], does also present eight criteria for a controlled medical terminology, and below we will discuss our model in the light of these.

3. Concept system

The concepts in a concept system is a construct of thoughts regarding important facts in the real world (the subject field) that we want to keep record of. We must be able to define them and to identify them in an unambiguous way. We designate the concepts by terms to be able to talk about and communicate them. One problem is that there often exists more then one designation for a concept (synonyms, alternative designations, local variants etc.). Sometimes the same term is used to designate different concepts i.e. homonyms. The structure in the concept system is expressed by different kinds of relations between the concepts (generic relations and semantic relations etc.). Generic relations (isa-relations) is a way to express specialization and generalization of concepts in hierarchies. A medical concept system cannot be arranged in a single strict hierarchy, since most concepts can have more than one superordinate concept. In other words it is a network structure.

There is often a need to be able to refer to or to translate into concepts in other terminological systems, classifications or nomenclatures. This can be handled by mapping or transformation. Semantic links (relations) can be used to represent even more terminological knowledge in a subject field:

e.g. infection which has_location = lung and has_cause bacteria = bacterial_pneumonia

3.1 The SpriTerm Conceptual Model

This is an abstract description of the Spriterm conceptual model. It does not cover security and historical information, neither details of data representation. The model is a result from several seminars held in Sweden and studies of international and national work in the area. The model is designed to meet both needs in developing a terminology and needs for using a terminology. It covers the requirement needed for an unambiguous and controlled terminology, but this has to be controlled via functions and rules. In this way the model can be used with different levels of ambition. We can use it as a structure holding just terms and verbal definitions as well as a structure able to represent information needed for a knowledge system with all knowledge represented by different kinds of relationships.

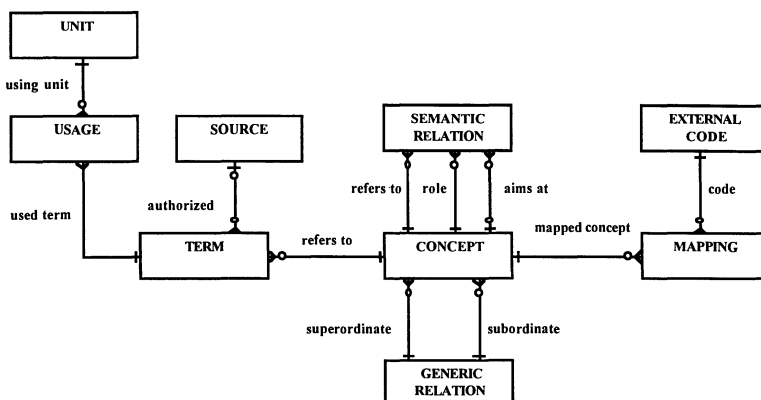


Fig 1. The SpriTerm Conceptual Model

Homonyms as well as synonyms are handled in a consistent way by qualifying TERM with the USAGE for a UNIT in the Swedish healthcare. A UNIT can be a medical speciality as well as a healthcare party. In this way a term is unambiguous in its context. References to other classifications, nomenclatures and terminologies are handled by MAPPING which also

qualifies the reference in terms of one to one, one to many etc. Generic relationships (isa) is represented by GENERIC RELATION able to cover a network structure, enabling a concept to have multiple parents. This way GENERIC RELATION can be used for implementing inheritance. Other relationships is covered by SEMANTIC RELATION which represents all other relationships including partitive relations as well as "caused by", "affects" etc. This structure can also be used to represent chapter structures of classifications as well as "search word" structures. In the model these types of relationships are also defined as concepts with associations to terms. Thus we can represent these relationships in the same way as other terms with usage, inheritance etc. Which enhances the flexibility of the model.

3.2. Description of the model semantics

This section will describe our model in the spirit of the recently issued MOSE document [1], with a meta-terminology for conceptual descriptions.

Terminological systems can to a large extent be described from their sets of concepts, semantic categories, with semantic links and associated categories that comprise valid associated domains. Codes will be entered manually and imported from different sources such as ICD-10 and NCSP, the Nordic Classification of Surgical Procedures.

According to this standard, the Spriterm system is compliant at level 1, since:

- Semantic categories are reasonably specified and include Administrative concepts, Diagnoses, Procedures.
- Semantic links are specified - we will initially use a limited set of links inspired by the UMLS metathesaurus and Read v3.
- Differentiating criteria are not listed in detail, but associated domains for semantic links will be specified, also in a manner similar to UMLS [6].

We have not devised any combinatorial rules among semantic categories and links, and therefore cannot claim compliance to level 2 of prENV 12264.

In comparison with the GALEN [4] concept representation, the Spriterm model is devised to be more pragmatic, and for the foreseeable future we may be forced to rely both on explicit textual definitions of concepts and on implied definitions based on semantic links.

4. Discussion about criteria's for a good terminological system for the healthcare sector.

Some of the criteria's following is mainly fulfilled through manual control or work while others are also depending upon logical structure. i.e. the conceptual model for the terminological system. In the following (M) means mainly manual control and (S) refers to structural issues (S,M) means both. All the criterias discussed below is supported by the Spriterm-model:

Clinical usefulness - i.e. the terminology must contain the right terms and concepts for use in different clinical situations. This means a terminology supported by the profession and accepted by the endusers. This also means that you need professional groups to get the right terms and concepts etc.(M).

Term triggered - the possibility to start with a term in natural language and to get the right unique concept directly. The user is not supposed to know which source to choose to find the right concept which for instance will be the case in a terminological system consisting of integrated sources (S).

Domain completeness - the system may not impose restrictions upon terminology e.g. by size, dimension or by coding system that restricts depth or breadth. The Spriterm model has no such restrictions (S).

Non redundancy - a concept may not exist more than once in the terminology (M).

Synonymy - support multiple non unique names for concept (S,M)

Non-vagueness - concepts in the terminology must be complete in meaning (M)

Definition by semantic links - define concepts by it's relationships to other concepts in the terminology (S,M)

Non ambiguity - concepts must have exactly one meaning else it is two different concepts (M)

Homonyms - the same term refers to more than one concept. This must be possible to handle in a correct way. In the Spriterm model it is determined by context (S)

Multiple classification - a concept may have more than one parent concept, i.e. network structure which the Spriterm model supports (S,M).

Consistency of use - concepts in multiple classes must have the same appearance in each context (M).

Explicit relationships - meaning of interconcept relationships must be clear (M).

Mapping - cross-references to other terminologies, classifications and nomenclatures (S,M).

Multilingual - possibilities to use or get terms in different languages (S,M).

Internationally harmonized - harmonized with international standard e.g. CEN etc. (M).

References:

- [1] CEN/TC251/PT2-003 prENV 12264: Medical informatics - Medical informatics - Categorical structures of systems of concepts - Model for representation of semantics. Bruxelles, 1996. CEN.
- [2] Cimino JJ, Clayton PD, Hripsack G, Johnson SB. Knowledge-based Approaches to the Maintenance of a Large controlled Medical Terminology. J Am Med Informatics Assoc. 1994; 1:35-50
- [3] Junfors A, Wallin S-B, Vad är ett termsystem för hälso- och sjukvården? En studie utifrån nationella och internationella erfarenheter. Stockholm, 1995. SPRI rapport
- [4] Rector AL, Zanstra P, Solomon D, GALEN Generalized Architecture for Languages Encyclopaedias and Nomenclatures in Medicine, Final Summary Report Dec 1994
- [5] NHS Centre for Coding and Classification, Read Codes File Structure Version 3.0 - Overview and Technical Description, jan 1995
- [6] Lindberg D, Humphreys B McCray A. The unified medical language system. Meth Inform Med 1993;32:281-291