Methods of Information in Medicine

N. F. de Keizer, A. Abu-Hanna

Understanding Terminological Systems II: Experience with Conceptual and Formal Representation of Structure

Department of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands

Abstract: This article describes the application of two popular conceptual and formal representation formalisms, as part of a framework for understanding terminological systems. A precise understanding of the structure of a terminological system is essential to assess existing terminological systems, to recognize patterns in various systems and to build new terminological systems. Our experience with the application of this framework to five well-known terminological systems is described.

Keywords: Terminological System, Conceptual Representation, Formalization

1. Introduction

For many decades various terminological systems have been developed with different domains and different structures, such as strict hierarchies or semantic nets describing concepts and their relationships. These terminological systems were designed for different purposes. Cimino et al. [1] and Campbell et al. [2] describe criteria concerning the essential conceptual features of an ideal terminological system. Chute et al. recently summarized and extended these criteria [3]. In spite of papers giving an overview of the strengths and weaknesses of terminological systems [2, 4, 5] it is still hard to gain insight into the merits and usability of existing systems because the structure and characteristics of terminological systems are often incompletely and ambiguously described. We feel there is a need for a framework for *understanding* terminological systems, a framework which is still lacking. To understand and compare existing terminological systems and to evaluate them for specific goals there is a need for at least two components: (1) a uniform terminology and typology to characterize terminological systems themselves [6] and; (2) a uniform re-presentation formalism to describe the *structure* of these systems which is essential for understanding and evaluating existing terminological systems or for the development of new ones.

The goal of this paper is to provide a representation formalism for representing the structure of terminological systems and to report our experience with its application for formalizing existing terminological systems. Essential in our representation formalism is that it is conceptual, viz. it supports communication between, for example, domain experts and engineers of the terminological system. It should also help to highlight weak spots in the design by supporting the comparison of various terminological system structures and the comparison between characteristics of the systems with those characteristics required. Therefore, complementary to the conceptual part, a formal counterpart (based on first order logic) is needed to enhance expressivity and disambiguity and to support consistency during development of new terminological systems and their maintenance and reuse.

This article describes our experience with formalizing five well-known terminological systems: ICD [7, 8]; NHS clinical terms [9, 10]; SNOMED [11, 12]; UMLS [13, 14]; and GALEN [15, 16] and the comparison of these systems with criteria believed important for an ideal terminological system [1, 2].

In Section 2 we describe the representation formalism which is based on Entity Relationship Diagrams (ERD) and First Order Logic (FOL). In Section 3 the relevant criteria of Cimino et al. [1] and Campbell et al. [2] are translated into this formalism. In Section 4 the (simplified) structure of some well-known terminological systems are conceptually and formally described and these structures are compared with the criteria of an ideal terminological system. An extensive description can be found in a technical report [17] obtainable from the authors. In Section 5 we discuss similarities and differences among, and implications for the usability of these terminological systems.

2. Uniform Representation Formalism

A uniform representation formalism supports the complete and unambiguous description of the structure of a terminological system enabling comparison of different terminological systems described by the same formalism. Important characteristics of a uniform representation formalism are: (1) conceptuality – i.e. it lends itself for human comprehension and communication, (2) adequate expressive power, and (3) non-ambiguity.

The Entity-Relationship (ER) formalism [18] is a simple formalism capable of expressing concepts, relationships between concepts and some cardinality constraints, and it scores well on these criteria. The simple diagrammatic notations of ER (ER Diagram) have contributed immensely to its popularity. The notations use in the static part of object-oriented formalisms, such as OMT and UML, could have also been used, because they basically express the same content. However, ERD may not always be adequate for expressing complex constraints. Hence, a more expressive instrument and a formal (based on mathematical notions) specification, is needed to complement it in order to avoid non-ambiguity and to capture complex constraints. Based on its expressive power and universality we have chosen (many sorted) First-Order-Logic (FOL). Our choice for ER with FOL means that descriptions in this formalism could easily be translated to and from other logic-based formalisms such as Ontolingua [19], conceptual graphs [20] and description logics [21] when their expressivity allows this. Other researchers have described the use of logic-based formalisms for the representation of medical data and pointed out that this is a pre-condition for automated reasoning [20-24]. Therefore, a formal and conceptual representation of the terminological system's structure, that is a meta-model of the medical concepts it includes, is essential in understanding the system and hence its usability.

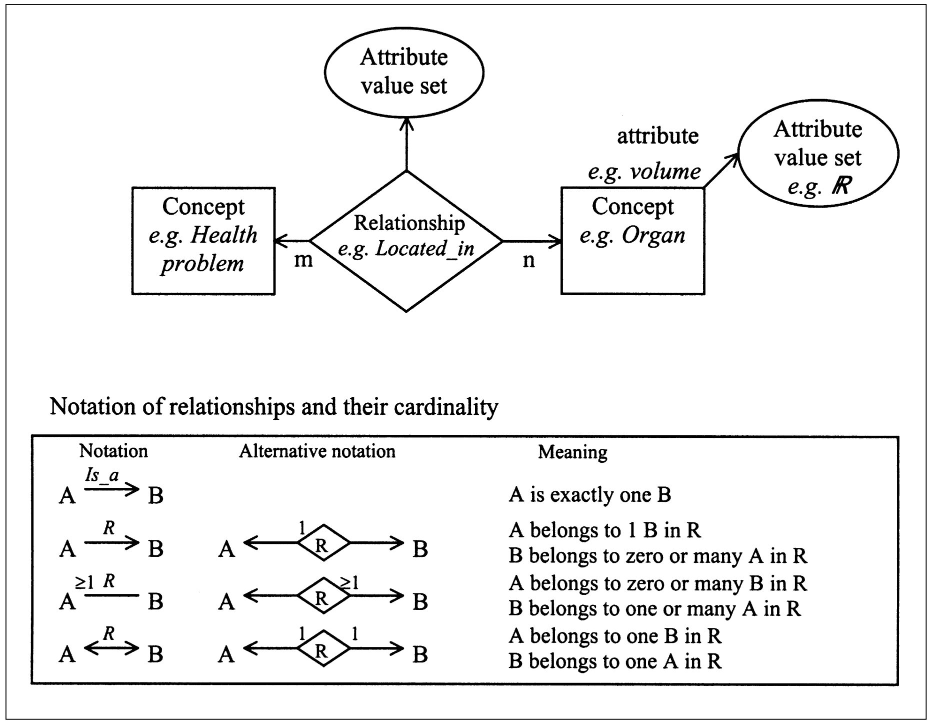


Fig. 1 ERD representation of concepts, attributes and relationships.

3. Representation of Criteria for Terminological Systems

Two categories of criteria of an ideal terminological system [1, 2] can be distinguished: criteria which concern the representation formalism itself, and criteria which concern the descriptions of the domain (the model) using the representation formalism. In this section both categories are described with the ER formalism and a FOL description. An extensive description can be found in a previous technical report [17].

3.1 Criteria Concerning the Representation Formalism

In this section we describe criteria for the representation formalism which form the basis for the domain criteria described in Section 3.2. As explained in [6] the building blocks of most conceptualizations and hence also of terminological systems are *concepts*, *attributes* and *relationships* between concepts; these are represented in ERD by a rectangle, an arrow connected with an ellipse, and a diamond connecting rectangles by arrows, respectively (Fig. 1).

Relationships between concepts can be distinguished in hierarchical relationships ("Is_a" and "Is_part_of" relationships) and non-hierarchical relationships (e.g. "caused_by"). When modeling the medical domain we represent an "Is_a" relationship between two different concepts as shown on the left side of Fig. 2, e.g. hepatitis *Is_a* liver disease. For the purpose of the description of models of terminological systems themselves we distinguish in

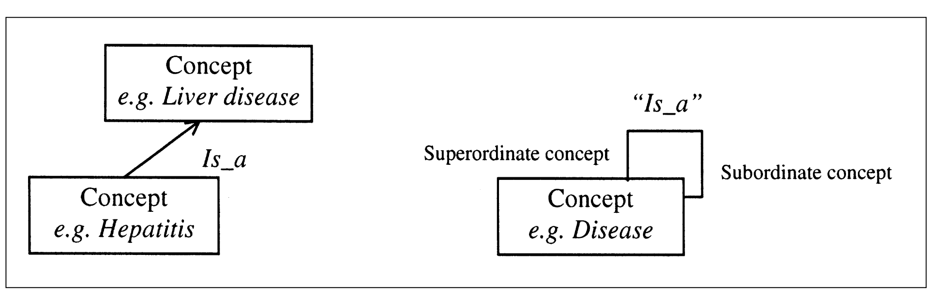


Fig. 2 ERD representation of hierarchical relationship normal (left) and at the meta-level (right).

this paper a meta-level “Is_a” relationship (right side of Fig. 2). In the meta-model the meta-concept has instances which are concepts at the domain level, e.g. “hepatitis” and “liver disease” are hierarchically related concepts and are instances of the meta-concept “disease”.

The rest of this section and Section 3.2 describe the criteria mentioned by Cimino et al. [1] and Campbell et al. [2]. A terminological system should enable the use of attributes to define or further specify concepts. *Relationships* between concepts should be *explicitly* represented by a label designating the meaning of the relationship, and constraints to restrict the interpretation of the relationship. Without explicitly representing relationships it is difficult to (automatically) interpret the meaning of a relationship, e.g., a relationship between “Health problem” and “Medication” can be interpreted as “Health problem *treated_by* Medication” or as “Health problem *is_side_effect_of* Medication”. *Composition rules* are grammatical rules, which define how and which concepts can be used to compose new concepts. With FOL one could easily specify the composites declaratively.

3.2 Domain Model Criteria

This section describes the criteria which concern the model of terminological systems.

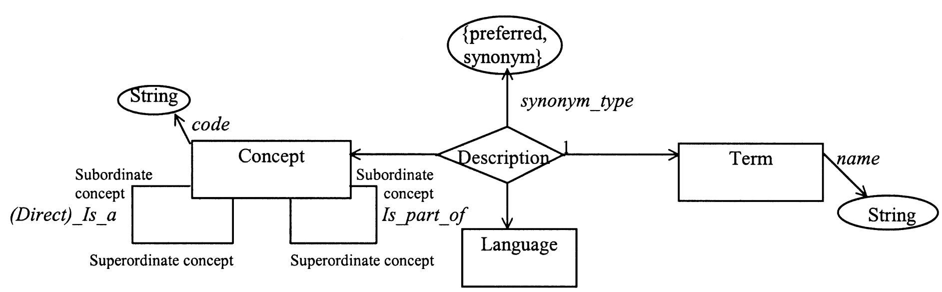
Domain completeness means that the terminological system should not be restricted in detail whether in depth or in breadth. Any constraint on the cardinality of the “Is_a” relationship or on the value of the (depth) level of the hierarchy, which reflects the chain of descendents of a concept, would hinder domain completeness (Fig. 3).

Synonyms and multi-lingual terms means that a unique concept may be designated by multiple terms in more than one language. Other model criteria concerning terms, concepts and their relationships are *non-ambiguity, non-vagueness* and *non-redundancy*. Non-vagueness prescribes that concepts must be complete in meaning, that is, refer to an object in the domain. Non-ambiguity prescribes that a concept refers to exactly one object in the domain. Non-redundancy prescribes that there should be a mechanism which prevents the existence of multiple different concept representations with the same meaning. These three criteria are not limited to the model of the terminological system, they are constraints on the meaning of concepts which ought to be considered by the knowledge engineer who develops the terminological system. To support *non-redundancy* in the model each concept has one preferred term and zero or more synonymous terms per language (see Fig. 3).

The next criterion *multiple classification* is represented by a hierarchical relationship in which a subordinate concept is related to as many superordinate concepts as required (see *Is_a* and *Is_part_of* relation in Fig. 3). For example “pneumococcal pneumonia” is a subordinate concept of the superordinate concept “lung diseases” as well as of the superordinate concept “infectious diseases”.

A code can be conceived as an attribute of a concept. Codes must be non-significant, i.e. *context free* hence not related to the meaning or the position of the concept in the hierarchy. Another criterion concerning codes is the possibility of *cross-mapping*, e.g., for administrative reasons. This can be observed or realized by an attribute “cross-mapping code” appearing at each concept. The cross-mapping code e.g., between ICD and a local terminological system, can be either manually or (semi)automatically derived.

The last criterion concerns the *use of definitions.* Some terminological systems have textual definitions which have to be interpreted by the human reader. If these definitions were formal, a computer could (at least partially) process them and use them for automated reasoning such as consistency checking [21] and knowledge acquisition in GAMES [24] and PROTÉGÉ [25].



- $\forall_{c \in \text{Concept}} \forall_{l \in \text{Language}} \exists!_{t \in \text{Term}}$ (Description(c,l,t) $\wedge$ synonym_type(<c,l,t>)=preferred))
  For every language there is exactly one preferred term per concept. $\exists!$y A means that exactly one “y” exists for which A is true.
- $\forall_{c1 \in \text{Concept}} \forall_{l \in \text{Language}} \forall_{t \in \text{Term}}$ (Description(c1,l,t) $\wedge$ synonym_type(<c1,l,t>)=preferred) $\rightarrow$
  $\neg\exists_{c2 \in \text{Concept}}$ (Description(c2,l,t) $\wedge$ synonym_type(<c2,l,t>)=preferred $\wedge$ c2$\neq$c1)
  For each language a term can be preferred for just one concept, i.e. different concepts cannot have the same preferred term.

To **avoid** multiple classification one could specify:

- $\forall_{c1,c2,c3 \in \text{Concept}}$ (Direct_is_a(c1,c2) $\wedge$ Direct_is_a(c1,c3) $\rightarrow$ c3 = c2)
  A subordinate concept c1 cannot have more than one superordinate concept

Fig. 3 ERD and FOL metadescription of non-redundancy, (multilingual) synonyms and multiple classification.

4. Description of Existing Terminological Systems

The conceptual and formal representation of terminological systems supports a better understanding of their structure. It helps to recognize the patterns in the designs of different terminological systems by a uniform view, which enables ascertainment of gaps and incompleteness in the terminological system.

In the first paper on this topic [6] we describe the typology and the coding scheme of the ICD-10 [8], NHS Clinical Terms [9, 10], SNOMED [11, 12], UMLS [13, 14] and GALEN [15, 16, 26]. We used ERD and FOL to describe the structure of these terminological systems and compare them with the criteria described in the previous section. For brevity, in this section we only represent the ERD and FOL description of ICD-10 and UMLS and only the essential information in FOL which is not represented in the ERD. An extensive conceptual and formal description of all systems is given in [17].

4.1 ICD-10

A conceptual and formal model of the ICD-10 is presented in Fig. 4.

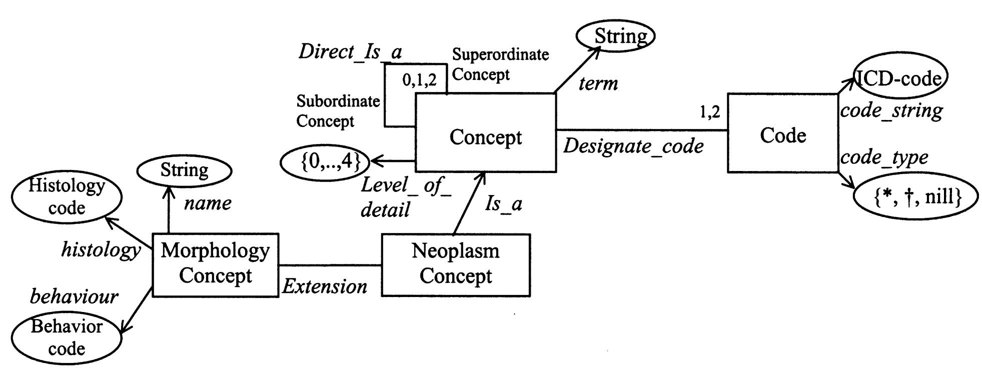
Although Fig. 4 shows explicit relationships to clarify the model of ICD-10, in reality the ICD-10 does not contain explicit relationships. For neoplasm concepts there are some term compositions. A coded nomenclature for morphology of neoplasms is part of the system. Each concept in the ICD-10 chapter 2 “Neoplasms” can be extended with a morphology concept and code, which consists of a histology code and a behavior code, e.g., “3=malignant, primary site”. For non-neoplasm concepts there are no attributes and no composition rules to compose new complex concepts.

Domain completeness of ICD-10 is restricted to 4 levels of depth. Comparing Figures 3 and 4 shows that there is no real distinction between concepts and terms, hence synonyms, multilingual terms and non-redundancy are not supported. From the comparison between Figures 3 and 4 we also conclude that multiple classification is restricted in ICD-10 by the cardinality: concepts have 0, 1 or 2 parents. Each concept is at least designated by one unique code and at most by two unique codes: one dagger code related to the etiology and one asterisk code related to the location of the diagnostic term. Definitions, beyond the implicit interpretation that a subordinate concept is a more specific form of the superordinate concept, are lacking.

4.2 NHS Clinical Terms

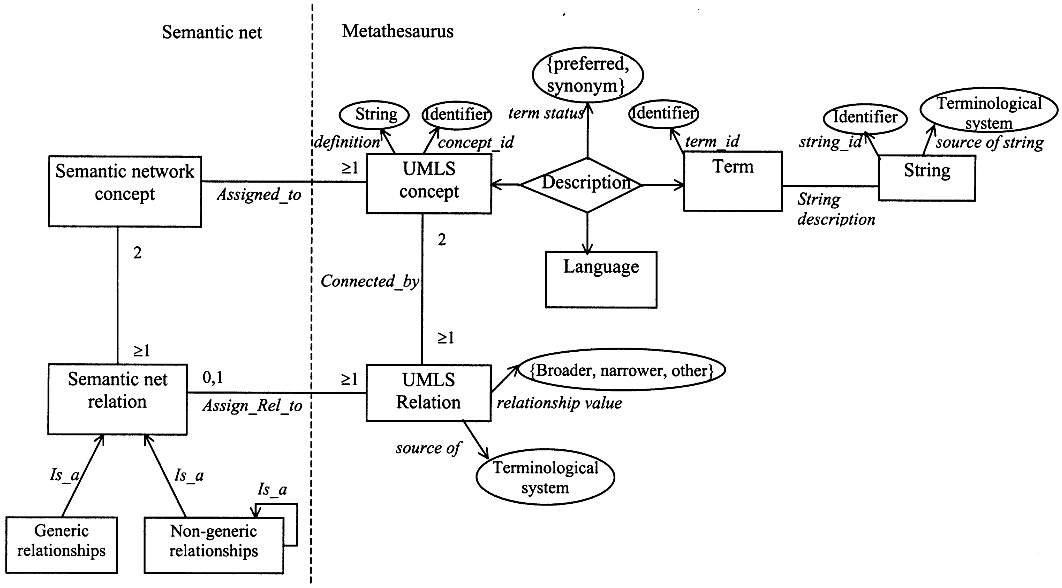
The NHS clinical terms, formerly known as the Read Clinical Classification forms a classification of medical concepts representing many domains such as diseases, signs, procedures, etc. Each of these subdomains contains concepts related by generic relationships. This system views partitive relationships as generic relationships by introducing *structure* concepts. For example, the subordinate concept “aortic arch” is part of the superordinate concept “thoracic aorta”, but instead of a direct partitive link this concept is generically linked to the *structure* concept “thoracic aorta structure”.

Although relationships are not formally made explicit, during the qualification of concepts, implicit relationships are used in the lookup tables (templates) to define combinations of concepts, attributes and attribute values in a controlled way [27]. The relationship between a concept and an attribute has a status: **Q**ualifier if the attribute might supply extra detail which a user might choose to further describe a concept, e.g., “course of illness: (acute/chronic)” to qualify “heart failure”; **A**tom if the attribute is an intrinsic characteristic of a concept, e.g., “site: cardiac structure” of heart failure. In our terminology this is part of a defini-



- $\forall_{c \in Concept}$ Level_of_detail(c) ≤ 4
  The number of levels of Direct_Is_A (detail in depth) is restricted to four.
- $\forall_{c \in Concept} \forall_{cd \in Code}$ Designate_code(c,cd) $\wedge$ $|\{<x,cd> \in$ Designate_code$|x=c\}| = 1 \rightarrow$ codetype(cd)=nill
  For each concept which is designated to exactly one code, that code has no specific type (nill)
- $\forall_{c \in concept} \forall_{cd1,cd2 \in code}$ Designate_code(c,cd1) $\wedge$ Designate_code(c,cd2) $\wedge$ cd1$\neq$cd2 $\rightarrow$
  ((code_type(cd1)= “*” $\wedge$ code_type(cd2)= “†”) $\vee$ (code_type(cd1)= “†” $\wedge$ code_type(cd2)= “*”))
  If a concept is designated by two different codes, one of the codes has the type “*” and the other code has type “†”

Fig. 4 ERD and FOL representation of ICD-10.



- $\forall_{c \in UMLSconcept} \forall_{l \in Language} \exists!_{t \in Term}$ (Description(c,l,t) $\wedge$ synonym_type($<c,l,t>$)=preferred))
  For every language there is exactly one preferred term per concept. $\exists!y$ A means that exactly one "y" exists for which A is true.
- $\forall_{c1 \in UMLSconcept} \forall_{l \in Language} \forall_{t \in Term}$ (Description(c1,l,t) $\wedge$ synonym_type($<c1,l,t>$)=preferred) $\rightarrow$
  $\neg\exists_{c2 \in Concept}$ (Description(c2,l,t) $\wedge$ synonym_type($<c2,l,t>$)=preferred $\wedge$ $c2 \neq c1$)
  For each language a term can be preferred for just one concept, i.e. different concepts cannot have the same preferred term.
- $\forall_{c \in UMLS\ concept} \exists_{s \in Semantic\ concept}$ Assigned_to (c,s)
  Each metathesaurus concept is assigned to at least one semantic concept in the semantic network

Fig. 5 ERD and FOL representation of the UMLS metathesaurus and semantic network.

tion. Template tables define which concepts can be modified by which attributes, but lack explicit constraints e.g., on combinations of attributes. Furthermore, existing atoms are not consistently used to describe the intension of their corresponding concept. Hence, definitions are unnecessarily partially specified.

NHS Clinical Terms has no limitations to domain completeness and multiple classification. Synonyms and unique codes are supported in the system. Each concept is designated by a description which consists of a unique code and a unique term (identifier) for the concept.

4.3 SNOMED

The structure of SNOMED consists of eleven modules (also called axes or dimensions), such as Topography, Disease and diagnosis, Procedures, etc., which can be conceived as distinct classifications. Concepts within one axis are related to each other using hierarchical relationships and concepts between different axes are also related by non-hierarchical relationships. An example of non-hierarchical relationships between terms are the relations between the disease "Tuberculosis" in the "Disease and diagnosis" and "Lung" in the topographical module, "Granuloma" in the morphology module, "M.tuberculosis" in the living organisms module and "Fever" in the function module. Since version II of SNOMED these relationships are explicit although a formal model lacks. Concepts are non-vague, that is they represent an object. Some terms, e.g., from the General modifier module, are vague but these are only used to compose new concepts. Concepts of the various modules/axes can be linked in order to compose new complex medical concepts, but SNOMED has not formalised any constraints on these compositions. SNOMED has no limitations to domain completeness and multiple classification. Multi-lingual terms, non-redundancy and unique codes are not supported. A disease concept can often be described by a non-vague concept from the "Disease and diagnoses" module, e.g., DE-14800 Tuberculosis, but in some cases it is also possible to describe the same concept with a concatenation of different concepts, e.g. Lung + Granuloma + M.tuberculosis + Fever which also represents Tuberculosis. Although each concept in SNOMED is described by one or more descriptions, there is no distinction between preferred and synonymous terms. Partial definitions are only available for concatenated concepts, in that case the definition is the enumeration of concatenated characteristics. SNOMED RT, which is under development, seems to be addressing all above-mentioned deficiencies by using a formal model [12] based on description logics [21].

4.4 UMLS

The UMLS consists of the Metathesaurus, the Semantic network, the SPECIALIST lexicon and the Information Sources Map. For brevity we only conceptually and formally describe the metathesaurus and the semantic network, because these two together form the most comparable component with the terminological systems described in the rest of this article. The Metathesaurus provides information about concepts, terms, string-names and the relationships between them, drawn from established terminological systems such as ICD-9-CM/ICD-10, SNOMED and MeSH. As shown in Figure 5 the metathesaurus represents "broader", "narrower", "other", relationships between different concepts (and optionally the relationships defined in the semantic network). Many relationships are derived directly from source terminological systems. For example, the hierarchical relationships in MeSH or ICD are manually made explicit in UMLS as "Is_a" or "Part_of" relationships. The UMLS does not contain composition rules to compose new complex concepts.

Comparing Figures 3 and 5 shows that UMLS has no limitations on domain completeness, and that multiple classification is possible. It also shows that synonyms, multiple-lingual terms, non-redundancy and unique codes are supported in UMLS. The formal specification of Figure 5 shows that each concept in the UMLS is described by one preferred and possibly more synonymous terms which are in turn linked to multiple strings (plurals, etc.). Each concept has an attribute "definition", a textual definition which describes the meaning of the concept.

Each concept in the metathesaurus is assigned to the most (one of 132) specific semantic category available in the semantic network. The semantic network provides information about the set of basic semantic categories (also called semantic types): Physical objects (e.g., organisms), Conceptual entities (e.g., findings), Activities (e.g., behavior) and Phenomenons and processes (e.g., biological function) and their relationships. Via the "Is_a" link, relationships and attributes are inherited by the subordinates of the high level semantic category. By inheritance the relationship "process_of" between "Biological function" and "Organism" also exists between "Disease or Syndrome" (which *Is_a* "Biologic Function") and "Human" (which *Is_a* an "Organism"). The inheritance of relationships can be blocked in case the subordinates of a semantic category conflict with the relationship, e.g. "Mental or Behavioral dysfunction" is a "Biologic Function" which can be related to an "Organism" by the "process of" relationship. "Plants" is a subordinate of "Organisms" but cannot have a mental dysfunction, therefore this inheritance is blocked. Relationships between semantic categories do not necessarily apply to all metathesaurus concepts that have been assigned to those semantic categories. For example, the relationship *"evaluation_of"* exists between the semantic categories "Sign" and "Organism attribute". The metathesaurus concept "overweight", related to the semantic category "Sign", is an *evaluation of* the "Organism attribute" concept "body weight" but it is not an *evaluation of* the "Organism attribute" concept "body length". Inconsistencies between metathesaurus concepts cannot be blocked.

4.5 GALEN

GALEN is different from the above-mentioned terminological systems. Like UMLS and SNOMED RT, GALEN provides an explicit model of the domain but it also provides a flexible representation language. GALEN's goal is to formally describe and model the medical domain by which the interchangeability of electronic medical data of different data sources can be supported. The Concept Module utilizes the GALEN Representation and Integration Language (GRAIL), a formalism based on Description Logics, to represent and manipulate the Concept Reference (CORE) model. The Core model is an ontology currently comprising approximately 5,000 concepts and 1,000 explicit relationships. The GRAIL formalism allows developers of terminologies to create models containing these concepts and relationship, and to derive (automatically) new composed concepts. There are no restrictions on domain completeness. To guarantee non-redundant and sensible composed concepts, automated reasoning facilitated by reasoning services of Description Logics [21], is used. Relationships be-tween concepts can be: "sanctioning", which specify how concepts are allowed to be used in the formation of composites; or "descriptive", which specify intrinsic characteristics of a concept. A conceptual and formal representation of a general GRAIL model is described in the technical report related to this article [17]. Automated reasoning within GRAIL facilitates multiple classification. Models developed with GRAIL are language independent and therefore the model of concepts is separated from the (synonymous) terms used to designate them. The Multilin-gual Module manages the mapping of concepts to preferred and synonymous terms of different languages. The Code Conversion Module can be used as an inter-lingua and manages the mapping of concepts to and from existing coding systems.

5. Discussion

In this article we have described the second part of a framework for understanding terminological systems and our experiences with its application: the use of conceptual and formal representation formalisms for representing the structure of a terminological system. After applying the formalism to the required criteria of an ideal terminological system (Section 3), we applied the representation formalism to describe the structure of important terminological systems (Section 4). In our experience, formalization supports the comparison between the criteria on terminological systems (Figs. 1 to 3) and the structure of existing terminological systems (Figs. 4 and 5). Formalization resulted in a reference design that helped us to observe discrepancies in some terminological systems. The largest problems with the ICD, which we observed during this exercise, are the lack of explicit relationships and definitions, the lack of separating terms and concepts (and so the lack of synonyms), the lack of possibilities for composing new concepts and the restrictions of domain completeness (number of levels in depth restricted to 4). We think that especially (formal) definitions and composition rules are essential criteria for future terminological systems because, especially when expressed in a restricted form, they can facilitate automated reasoning such as consistency checking, classification [21] and knowledge acquisition (data entry) [24, 25].

The NHS Clinical Terms, SNOMED and the UMLS do better on most criteria, but composition rules and formal definitions are also missing or premature. SNOMED RT aims to address this deficiency but it is not yet operational. The UMLS uses a semantic network to structure concepts in the medical domain. Reasoning is not (optimally) supported in the UMLS because nodes and arcs are only intuitively defined by their labels on a high level of detail, and therefore sensibility control on metathesaurus concept level is not supported.

The GALEN project, not extensively operational yet, is an ambitious and promising project aiming at a formal description of the total medical domain. GALEN has the intention to adhere to the criteria mentioned in Section 3 but the realization of all these intentions is still under construction. Moreover, the restrictions of the description logics on which GALEN is based, improves reasoning with the system, but constrains its expressivity. But even the somewhat restricted expressiveness of GALEN and its orientation towards the total medical domain still implies that a great effort should go into the syntax and grammatical rules to guarantee sensibly composed concepts. We intend a thorough and practical evaluation of the CORE model and the GALEN applications in a collaborative research. Special attention will be paid to the evaluation of the expressiveness and the possibilities for automated reasoning in clinical settings.

We believe that the formalization of terminological systems helped us with a thorough understanding of their structure and merits. It was also the basis for the development of a new terminological system for intensive care [28]. In our opinion, formal representation formalisms constitute indispensable tools for serious terminological system developers. The representation of the structure of a terminological system in a conceptual and formal way has more advantages next to merely *understanding* the terminological system [20, 22-24, 29]. A conceptual and formal representation of a structure of a system supports the communication about what the system means and it supports development of new systems by finding the "desired patterns" (criteria of Section 3) in the design and by building new designs based on desired ones. Furthermore, by making knowledge of the underlying domain explicit with formal specifications, these specifications can be used in a knowledge acquisition tool, such as GAMES [24] and PROTÉGÉ [25], to support inference of new knowledge and to support consistency checks within the terminological system. These are important tools for the management of the terminological system.

Acknowledgement

We thank Ronald Cornet for his valuable discussions and helpful remarks on an earlier draft of this manuscript.

REFERENCES

1. Cimino J, Clayton P, Hripcsak G, Johnson S. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Informatics Assoc 1994; 1: 35-50.
2. Campbell J, Carpenter P, Sneiderman C, et al. Phase II evaluation of clinical coding schemes: completeness, definitions and clarity. J Am Med Inform Assoc 1997; 4: 238-51.
3. Chute C, Cohn S, Campbell J, et al. A framework for comprehensive Health Terminology Systems in the United States: Development guidelines, criteria for selection and public policy implications. J Am Med Inform Assoc 1998; 5 (6): 503-10.
4. Cimino J. Coding Systems in Health Care. Yearbook of Medical Informatics 1995: 71-85.
5. Klimczak J, Hahn A, Sievert M, Petroski G, Hewett J. Comparing clinical vocabularies using coding system fidelity. In: Gardner R, ed. Annual Symposium on Computer Applications in Medical Care. New Orleans: Hanley & Belfus 1995.
6. de Keizer N, Abu-Hanna A, Zwetsloot-Schonk J. Understanding terminological systems I: terminology and typology. Method Inform Med 2000; 39: 16-21.
7. International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 9th revision. WHO 1977.
8. International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 10th revision. WHO 1993.
9. Robinson D, Comp D, Schulz E, Brown P, Price C. Updating the Read Codes: User-interactive Maintenance of a Dynamic Clinical Vocabulary. J Am Med Inform Assoc 1997; 4 (6): 465-72.
10. Schulz E, Price C, Brown P. Symbolic Anatomic Knowledge Representation in the Read Codes Version 3: Structure and Application. J Am Med Inform Assoc 1997; 4: 38-48.
11. Rothwell D. SNOMED-Based knowledge representation. Method Inform Med 1995; 34: 209-13.
12. Spackman K, Campbell K. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Am Med Inform Assoc annual symposium, 1998: 740-4.
13. UMLS Knowledge sources – Documentation. National Library of Medicine 1997:
14. Lindberg D, Humphreys B, Mc Cray A. The Unified Medical Language System. Method Inform Med 1993; 34: 281-91.
15. Rector A, Solomon W, Nowlan W, Rush T, Zanstra P, Claassen W. A Terminology Server for medical language and medical information systems. Method Inform Med 1995; 34: 147-57.
16. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The Grail concept modelling language for medical terminology. Artif Intell 1997; 9: 139-71.
17. de Keizer N, Abu-Hanna A. A framework for understanding Terminological Systems. Amsterdam: Academic Medical Center. Dep Medical Informatics 1999.
18. Chen P. The Entity-Relationship model; toward a unified view of data. ACM Trans on Database Systems 1966; n: vol 1:1.
19. Gruber T. A translation approach to portable ontology specifications. Knowledge Acquisition 1993; 5: 199-220.
20. Campbell K, Das A, Musen M. A logical foundation for representation of clinical data. J Am Med Informatics Assoc 1994; 1: 218-32.
21. Donini F, Lenzerini M, Nardi D et al. Reasoning in Description Logics. In: Brewka G, ed. Principles of knowledge representation and reasoning. CLSI publications, 1996: 193-238.
22. Abu-Hanna A, Jansweijer W. Modeling application domain knowledge using explicit conceptualization. IEEE-Expert 1994.
23. Gruber T. Towards principles for the design of ontologies used for knowledge sharing. Int J Hum Comp Stud 1995; 43: 907-28.
24. Heijst G, Falasconi S, Abu-Hanna A, Schreiber A, Stefanelli M. A case study in ontology library construction. Art Intell Med 1995; 7: 227-55.
25. Studer R, Eriksson H, Gennari J, Tu S, Fensel D, Musen M. Ontologies and the Configuration of Problem-solving Methods. In: Gains B, Musen M, eds. 10th Banff Knowledge Acquisition for Knowledge-

Based Systems Workshop. Banff, Canada 1996.
26. Rector A, Glowinski A, Nowlan W, et al. Medical-concept models and medical records: an approach based on GALEN and PEN&PAD. J Am Med Inform Assoc 1995; 2: 19-35.
27. Read Code File structure version 3.1. NHS Centre for Coding and Classification 1995.
28. Campbell K, Oliver D, Spackman K, Shortliffe K. Representing Thoughts, Words, and Things in the UMLS. J Am Med Inform Assoc 1998; 5 (5): 421-31.
29. Moorman P, van Ginneken A, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Yearbook of Medical Informatics 1995: 195-204.

Address of the authors:
Nicolette F. de Keizer,
Department of Medical Informatics, J2-256,
Academic Medical Center, P.O. Box 22660,
1100 DD Amsterdam, The Netherlands
E-mail: n.f.keizer@amc.uva.nl