

A Framework for Characterizing Terminological Systems

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

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

Summary

Objectives: The notion of a terminological system (TS) is complex due to the broad range of systems, applications, and clinical domains. A uniform approach to describe the characteristics of TSs is lacking. This impedes further understanding, applicability, mutual comparison and development of TSs. For these reasons we propose a terminological systems characterization framework.

Methods: Relevant issues pertaining to TSs and terminology servers have been extracted from literature describing requirements and functionality of TSs. From these issues, features have been distilled and further refined. A categorization has been developed to provide a convenient arrangement of these features.

Results: The framework distinguishes between application-dependent and application-independent features of TSs. Definitions are provided for measures of content coverage, which was identified as the only application-dependent feature. Application-independent features are categorized along two axes: their respective type of TS and the particular element within that system, i.e. the formalism, the content, or the functionality. For each feature we provide an explicit question, the answer to which yields a feature value. The framework has been applied to SNOMED CT and the CLUE browser.

Conclusions: We present and apply a framework to support a feature-based characterization of terminological systems. Standardized methods for content coverage studies reduce the effort of assessing the applicability of a TS for a specific clinical setting. A two-axial categorization provides a convenient arrangement of the large number of application-independent features. Application of the framework increases comparability of terminological systems. This framework may also help TS developers determine how their system can be improved.

Keywords

Terminological systems, evaluation, methods, definitions

Methods Inf Med 2006; 45: 253–66

1. Introduction

Changes in health care organization and technological development have resulted in different health care information systems. This has led to an evolution of medical terminological systems aimed at satisfying the demands for re-use and faithful transmission of data and computer-based management of semantics. In accordance to [1] we define a terminological system as “a model of concepts and relationships together with the terms pertaining to them”.

Rossi Mori et al. [2] describe the evolution of terminological systems in terms of three generations. First-generation systems, e.g. the ICD-family [3] and the Medical Subject Headings (MeSH) [4], are characterized by a fixed organization (typically hierarchical) and a simple representation such as a systematic list that is alphabetically indexed. Second-generation systems, such as the medical dictionary for regulatory activities MedDRA [5], LOINC [6] and SNOMED International [7], have a dynamic organization (i.e. provide multiple hierarchies) and are compositional, combining the simple list representation of concepts with a knowledge base to define and extend these concepts. Third-generation systems, e.g. SNOMED CT [8], GALEN [9], Gene Ontology (GO) [10] and the Foundational Model of Anatomy (FMA) [11], are based on formal models providing symbols denoting concepts and a set of formal rules to manipulate them. Throughout these generations, terminological systems have developed from single-purpose, inextensible systems to extensible multi-purpose systems. The range of domains that terminological systems cover is broad, as is indicated by the examples above. It covers among others patient data, anatomy, drugs, genomics, and medical literature. The in-

crease of terminological systems both in number and size is demonstrated by the growth of the UMLS Metathesaurus which integrates a large number of terminological systems [12]. The 2004AC Metathesaurus contains information about over 1 million biomedical concepts and 4.3 million concept names (i.e. terms) from more than 100 terminological systems^a.

Due to the multiplicity and dynamics of terminological systems, a need for understanding their characteristics has emerged. Based on a review of literature and relevant standards, a typology of these systems is defined [1], which is summarized in Table 1. Each terminological system is a terminology, i.e. a list of terms denoting concepts in a domain, with possibly additional characteristics, e.g. it is also a vocabulary when the system includes definitions of the concepts of the terminology. Based on this typology, (recent versions of) the International Classification of Diseases (ICD) can be typified as not only a classification, but also a thesaurus, terminology and a coding system. As another example, SNOMED CT (Systematized Nomenclature of Medicine) is not only a nomenclature, but can also be typified as all of the system types mentioned in Table 1.

This typology provides a first means for categorizing terminological systems. Other approaches have been advocated that distinguish terminological systems for example by their prototypical use. In [13], systems for recording detailed patient data are referred to as nomenclatures, whereas systems used for statistical purposes are referred to as (statistical) classifications.

The fields of application of medical terminological systems have expanded over

^a <http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html> (last visited February 14, 2005)

Table 1 Overview of types of terminological systems, as defined in [1]. Each terminological system is a terminology and possibly one or more of the following: thesaurus, classification, vocabulary, nomenclature, and/or coding system.

Type of terminological system	Distinctive characteristic
Terminology	List of terms referring to concepts in a defined particular domain.
Thesaurus	Terms are ordered e.g. alphabetically Concepts are described by more than one (synonymous) term.
Vocabulary	Concepts have definitions, either formally or in free text
Nomenclature	A set of rules for composing new complex concepts or the terminological system resulting from this set of composition rules
Classification	Concepts are arranged using generic (is_a) relationships
Coding system	Codes designate concepts

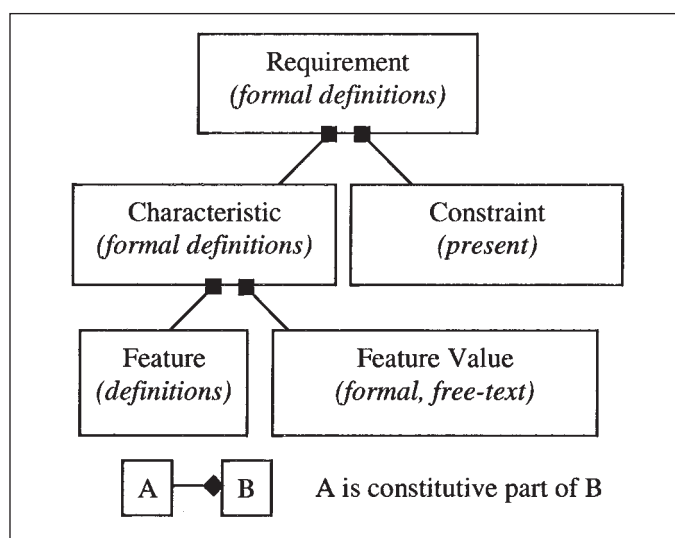


Fig. 1 Schematic presentation of the notions used in the literature for various issues of terminological systems. Examples are given in italics.

the years. One of the first terminological systems, the ICD, was developed in order that “the medical terms reported by physicians, medical examiners, and coroners on death certificates can be grouped together for statistical purposes”. Contemporary terminological systems enable a much broader use; the NHS Information Authority sketches a spectrum of applications for terminological systems: Documentation in the EPR/EHR; Decision support; Clinical audit; Reporting; Summaries; Administrative and management information; Epidemiology; Billing; and Resource management^b.

Given this broad range of possible applications and the large number of highly different terminological systems, it is hard to compare TSs and to determine which terminological system(s) can fulfil specific needs. There is no generic solution satisfying all needs, and the usefulness of a specific terminological system may differ from one situation to another. The aim of this paper is to provide a framework to describe features of TSs. This framework can support comparison between terminological systems, assessing the fulfilment of requirements and development of a terminological system. In order to explicitly distinguish the various notions used throughout this paper, Figure 1 gives a schematic presentation of the notions used in this paper. Figure 2 provides a simplified representation of the process of determining the applicability of a

terminological system to meet application-specific requirements. A requirement consists of a characteristic and its constraint(s). Characteristics of a TS can be made explicit by features and their values. In the framework presented in this paper, features (also called attributes, e.g. “number of concepts”, “underlying formalism”) are explicitly distinguished from feature values (a.k.a. attribute values, e.g. “1038 concepts”, “frames representation”). To determine the degree to which a requirement is satisfied, the feature values of terminological systems are matched with application-specific requirements. This requires the existence of a well-defined and well-described set of features and feature values of terminological systems. However, such a description of features and their values does not currently exist, which hampers the assessment of the applicability of a TS and the comparability of terminological systems. As shown in Figure 2 our approach for extracting these features and their values consists of two steps. In the first step, requirements relevant to terminological systems are identified. In the second step possible features are derived and questions are formulated to obtain feature values of a TS. This enables the characterization of terminological systems in a structured way, thus providing insight into the similarities of and differences between various terminological systems. Thereby it increases the comparability of terminological systems and the support for assessment of their applicability. As the total number of features may become very large, the focus in this paper will be on the intrinsic features of terminological systems. Hence licensing issues or organizational topics such as maintenance or versioning will not be discussed.

It has often been pointed out (e.g. in [14, 15]) that desired characteristics of and criteria for terminological systems may vary with their intended usage. In [16], this application-dependence is mentioned as the first barrier to evaluation of terminological systems. Contrary to desired requirements (e.g. “concepts must be designated by Dutch terms”), values of features of terminological systems are almost exclusively application-independent (e.g. “concepts are designated by terms in English and French”).

^b http://www.nhsia.nhs.uk/snomed/pages/events/snomed_update/text5.html (last visited February 14, 2005)

The only application-dependent feature identified is “content coverage”, a quintessential feature of terminological systems. Our framework explicitly distinguishes between (application-dependent) content coverage and application-independent features, as shown in Figure 2.

This paper is organized as follows. An overview of related research, and the background for the framework is presented in Section 2. Section 3 provides the process of developing the framework for characterizing terminological systems. Section 4 concerns the application-dependent feature “content coverage” and summarizes methods to evaluate the content of a terminological system. Furthermore, Section 4 presents a categorization of features formulated as explicit questions for the application-independent description of terminological systems. Section 5 presents the application of the framework to the SNOMED CT terminological system^c, which is receiving increasing attention. Section 6 discusses the merits and limitations of this framework by looking at various applications of the framework. This section also relates the proposed framework to the literature and addresses issues that require further research. Section 7 concludes this paper.

2. Background

Although first-generation terminological systems were developed in a paper-based era, this does not mean that these systems are useless in today’s computerized environment. Each of the three generations of terminological systems does have its advantages and disadvantages in terms of use, maintenance and costs. Therefore it is important to understand the features that characterize these systems and evaluate them with regard to the requirements of their potential users. Fortunately the topics of standardization and understanding of terminological systems are getting increasing attention. This has resulted in various

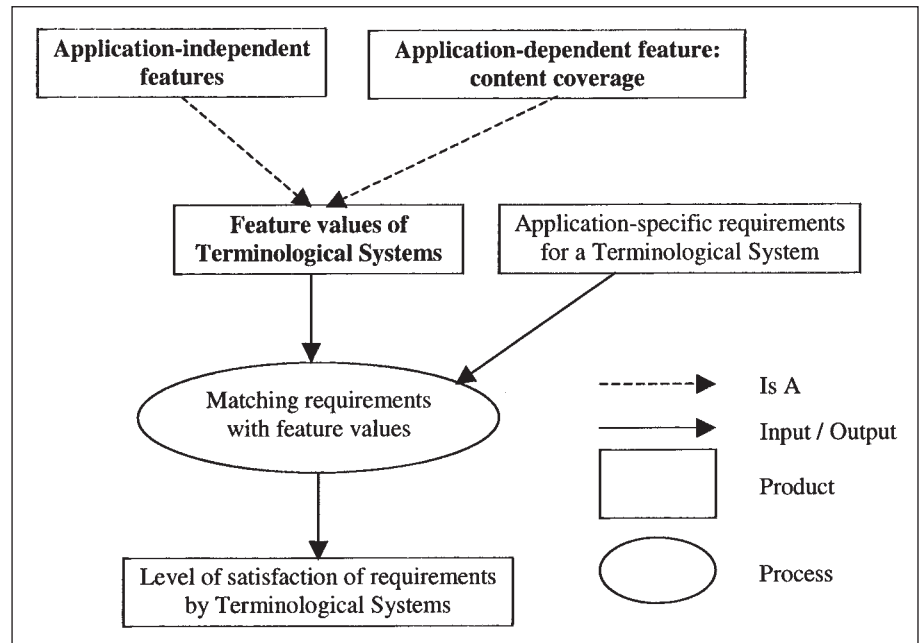


Fig. 2 Schematic representation of using features of terminological systems for determining the satisfaction of requirements for a specific application

publications that address description and evaluation of terminological systems and terminology servers [14-18]. We define terminology servers as “software modules that provide functionality for navigation, manipulation and/or modification of a terminological system by means of a (standardized) application-programming interface”. These modules are often closely tied to a specific terminological system and therefore we include their functionality in our framework.

A number of publications have paid attention to describing requirements for terminological systems. Among these publications is [14], specifying twelve desiderata that were distilled (mainly) from literature from the 1990s. The desiderata (such as “concept orientation”, “polyhierarchy” and “formal definitions”) are proposed as a checklist to address the requirements of intended users of second and third-generation terminological systems. The desiderata state the characteristics a terminological system should possess, but do not pay attention to methods for measuring these characteristics, their significance or interdependence.

The “Standards Specification for Quality Indicators for Controlled Health

Vocabularies” [15] is a further step towards structured specification of terminological systems. Largely based on [14], it distinguishes “general information and characteristics” and “characteristics describing the structure” of the terminology model. Furthermore, it describes characteristics influencing maintenance, and characteristics and measures for the evaluation of terminological systems.

In the USA in 2003, the National Committee on Vital and Health Statistics Subcommittee on Standards and Security has made an inventory of about 40 terminological systems to arrive at national terminology standards for Patient Medical Record Information [17]. This inventory was based on a questionnaire that contained between 40 and 100 (depending on the level of detail considered) questions regarding a large number of characteristics of terminological systems (and their developers). This questionnaire intends to use independent, continuous measures of well-defined characteristics without paying attention to their significance and interdependence. As such, it is the first effort known to the authors to describe a large number of terminological systems in a structured manner.

^c <http://www.snomed.org/snomedct/index.html> (last visited February 14, 2005)

The Object Management Group (OMG) has taken a functionality-oriented approach in the Lexicon Query Service Specification (LQS) [18], which defines “methods for accessing the content of medical terminology systems”. Rather than defining characteristics of terminological systems, it provides a reference of functions that terminology *servers* should offer. As many functions depend on characteristics of the underlying terminological system, the functionality also provides insight into requirements on a terminological system.

Other recent research focuses on the ontological correctness of the contents of terminological systems [19, 20]. These papers provide an analysis of the interactions between ontological and epistemological components of terminological systems, and on the distinction between classes and concepts. Analyses in [20] aim at determining among others: terms containing classification criteria, and terms reflecting detectability, modality, uncertainty, and vagueness. In [19] a discussion is provided on the use of classes and concepts in terminological systems, where classes do indicate naturally delimited sets (e.g. fracture, breast cancer), as opposed to concepts, that provide artificially constructed sets (e.g. fracture without intracranial injury). Both papers demonstrate the need for in-depth study of the contents of terminological systems and the need for developing methods to guard their ontological correctness.

Recently, a framework comparable to the one we present in this paper, has been developed [21, 22]. The distinction between the frameworks is that the work of Supekar focuses on formal ontologies in the context of the semantic web, whereas we aim at providing a more generic description, not only of systems based on formal representation (i.e. ontologies), but also on traditional terminological systems. Moreover, our framework is restricted to terminological systems in the domain of health care.

This overview shows that various efforts have been made towards description or (means for) evaluation of terminological systems. This paper builds on these efforts, by proposing a general framework that explicitly distinguishes features and feature values, categorizes these features, and

suggests methods to determine the feature values in an objective and reproducible manner.

3. Process of Formulation of the Framework

The framework presented in this paper consists of features that we have extracted from the literature and then categorized, in order to show the interdependence and significance of the features. Furthermore, features have been refined when appropriate, and methods to determine feature values are provided where applicable.

3.1 Collection of Features

The work mentioned in Section 2 [14, 15, 17, 18] forms our starting point for the literature that was consulted. This is due to the generic approach of these papers and because of their focus on requirements for terminological systems. From these papers, issues relevant for characterizing terminological systems were extracted. These issues are either *desiderata* (required characteristics), or generic characteristics (ways of describing terminological systems). Based on the issues found, we have derived relevant features.

The value of the feature “content coverage” (the extent to which the concepts and terms used in the terminological system cover the domain) is highly application-dependent. The content coverage depends on the intended application and domain of use; hence this cannot be determined independently of the application of a terminological system. In order to make outcomes of content coverage studies comparable, agreed-upon methods for assessing content coverage are useful. Such methods are described in Section 4.1.

In contrast to “content coverage”, all other characteristics of terminological systems can be assessed independently of an application. For example, whether a system provides “synonyms” is independent of the user of the system, e.g. physicians or researchers in internal medicine, surgery or

primary care. The application-independent features are organized according to the categorization presented in the next subsection.

3.2 Categorization of Application-independent Features

To provide further structure for the set of application-independent features, we categorize these according to two axes: the elements of terminological systems and servers, as described below, and the type of terminological system (terminology, thesaurus, vocabulary, nomenclature, classification, coding system, as depicted in Table 1).

3.2.1 Elements of Terminological Systems and Servers

The “elements of terminological systems and servers” axis consists of “formalism”, “content (domain knowledge)”, and “functionality” [23]. We take this issue of functionality of terminology servers into account as many contemporary terminological systems are packaged with some “default” services, and because the use of a terminological system in a computerized environment is commonplace. Moreover, issues mentioned in literature often involve both systems and servers.

Within a terminological system, one can distinguish the *domain knowledge* and the *formalism* that is used to represent the domain knowledge. The formalism (e.g. frame-based representation, entity-relationship modeling, or description logic) is fully separated from the represented domain knowledge. Others further subdivide concepts and relations into, for example, “categorical structure” (a meta-model of concept classes and their relations) and “system of concepts” (the set of concepts of the specific domain) [24], or “Top-level Ontology”, and “Domain Ontology” [25]. As such a distinction is relevant for concepts and relations, but not for other content, such as the terms or codes in a terminological system; we do not further subdivide domain knowledge.

Table 2 Illustration of the process of extracting features from the literature and categorizing them, applied to “desiderata for controlled medical vocabularies” from [14]. The first two columns provide quotations from [14], the additional remarks and categorization in the last two columns are provided by the authors.

Desideratum	Mentioned Issues	Additional remarks	Categories
Content	<ul style="list-style-type: none"> ● Add terms as they are encountered ● Compositional extensibility ● Formal methodology for expanding content ● Methods for recognizing and filling gaps in content 	<ul style="list-style-type: none"> ● In a wider definition, content refers to terms, concepts, relations, composition rules ● No restrictions may be put on the breadth or depth of the taxonomy ● Compositional extensibility implies that a system is a nomenclature ● Maintenance must be based on formal methods 	formalism content terminology classification nomenclature coding
Concept Orientation	<ul style="list-style-type: none"> ● Nonvagueness ● Nonambiguity ● Nonredundancy, in the context of pre-coordinated concepts 	<ul style="list-style-type: none"> ● These issues need to be asserted during modeling ● Nonvagueness is very hard to evaluate ● Nonambiguity and Nonredundancy can partly be evaluated when (formal) definitions are available 	content vocabulary nomenclature
Recognize Redundancy	<ul style="list-style-type: none"> ● Provide a mechanism by which redundancy can be recognized in the context of post-coordination 	<ul style="list-style-type: none"> ● This is functionality that depends on the representation 	formalism function vocabulary nomenclature
Formal Definitions	<ul style="list-style-type: none"> ● Expressed using e.g.: frames, semantic networks, classification operators, categorical structures, conceptual graphs 	<ul style="list-style-type: none"> ● The formalism needs to be explicitly described; e.g. supported structures and their semantics ● Most notably: Description Logics 	formalism vocabulary
Representing Context	<ul style="list-style-type: none"> ● Coping with contexts may be easier if such contexts are modeled in the vocabulary 	<ul style="list-style-type: none"> ● This is not necessarily part of the vocabulary, but loosely related to it 	formalism vocabulary
Polyhierarchy	<ul style="list-style-type: none"> ● Allow multiple hierarchies to coexist 	<ul style="list-style-type: none"> ● Concepts can have multiple parents ● Other non-taxonomic hierarchies (e.g. paronymy) must be possible 	formalism classification vocabulary
Concept Permanence	<ul style="list-style-type: none"> ● The meaning of a concept is inviolate 	<ul style="list-style-type: none"> ● This has to be asserted during modeling ● Concept deletion is not allowed, a mechanism is required for marking a concept “obsolete” ● It is unclear how to deal with concepts for which the set of subordinate concepts has changed 	formalism content classification vocabulary
Evolve Gracefully	<ul style="list-style-type: none"> ● Give a clear, detailed description of what changes occur and why 	<ul style="list-style-type: none"> ● This is a maintenance issue, not an intrinsic feature 	-
Multiple Granularities	<ul style="list-style-type: none"> ● The more macroscopic the level of discourse, the coarser the granularity of the concepts; hence vocabularies be capable of handling both fine-grained and general concepts 	<ul style="list-style-type: none"> ● Like “content”, no restrictions may be put on the breadth or depth of the taxonomy 	formalism content classification vocabulary
Multiple Consistent Views	<ul style="list-style-type: none"> ● An application may restrict coding to coarse-grained concepts, hide intermediate classes or limit the user to a single, strict hierarchy 	<ul style="list-style-type: none"> ● This requires representation of “relevance” for various domains ● It is arguable whether this is part of the terminological system 	function vocabulary
Nonsemantic Concept Identifier	<ul style="list-style-type: none"> ● Concepts must have a unique identifier, free of hierarchical or other implicit meaning 	<ul style="list-style-type: none"> ● Using a random identifier is required 	formalism coding
Reject “N.E.C.”	<ul style="list-style-type: none"> ● Definition can only be based on knowledge of the rest of concepts in the vocabulary, leading to “semantic drift” 	<ul style="list-style-type: none"> ● This can be regarded a versioning problem, but poses constraints on domain knowledge 	content classification vocabulary

- *Formalism*-related features are those that relate to the formalisms underlying the representation of terminological knowledge. For example, whether the formalism of the system allows the expression of a poly-hierarchy, or whether the formalism restricts the maximum granularity.
- *Content (domain knowledge)*-related features describe the actual content of (a specific version/release of) a system. Examples thereof are the number of concepts, the average number of parent concepts (to measure use of poly-hierarchical definitions), and the covered clinical domains. Note that these characteristics present an overview of the content. Statements about e.g. the completeness of the terminological system’s content are part of the application-dependent characterization of terminological systems.

Table 3 Definitions for various measures for content coverage

<i>Concept coverage:</i> the extent to which the concepts within a subset, representative for the domain of interest, can be represented by the concepts within the terminological system.
<i>Concept token coverage:</i> concept coverage using a subset in which each concept may occur more than once, indicating the occurrence of that concept in practice.
<i>Concept type coverage:</i> concept coverage using a subset in which each concept occurs at most once.
<i>Post-coordinated concept coverage:</i> the extent to which the concepts within a representative subset can be represented by the concepts (either pre-existing or created with use of composition rules) within the terminological system.
<i>Term coverage:</i> the extent to which the terms within a representative subset exist in the terminological systems' content, provided that the terms relate to concepts that are present in the terminological system.
<i>Term token coverage:</i> term coverage using a subset in which each term may occur more than once, indicating the occurrence of that term in practice.
<i>Term type coverage:</i> term coverage using a subset in which each term occurs at most once.

- *Function*-related features describe a terminology server in terms of the provided functionality, e.g. retrieval of descendant concepts, or translating a term from one language to another. Ideally, a terminology server is separated from a terminological system, so that a server can be used with more than one system, and likewise, a system can be addressed by more than one server.

The two axes described above result in a 3 by 6 grid in which application-independent features are placed. This provides explicit and comprehensive clusters. We will use this grid in Section 4.2.

3.3 Refinement of Features

The process of placing application-independent features into the above-mentioned grid frequently required further refinement of features, as placement of conceived features was non-trivial. We also defined and categorized additional features, similar to the ones found in the literature but not mentioned as such. To illustrate how we performed the process of feature extraction, refinement and categorization, Table 2 shows how the desiderata from [14] have been processed. Criteria from other literature mentioned have been processed in the same way, but this is not represented in Table 2. The two columns on the left describe the issues mentioned for each desideratum. The two columns on the right present additional

remarks to these issues, plus the relevant categories of the features from the two axes. For example, in [14] the desideratum "Recognize Redundancy" (as expressed in column 1) is defined and the accompanying text (summarized in column 2) mentions "As vocabularies evolve, gracefully or not, they will begin to include this kind of redundancy [i.e. multiple ways to code a concept]. Rather than pretend it does not happen, we should embrace the diversity it represents while, at the same time, provide a mechanism by which we can recognize redundancy and perhaps render it transparent." The third column provides a summary of the analyses of the issues mentioned in the literature, which lead to the applicable categories of the framework (i.e. formalism and functionality of both the vocabulary and nomenclature).

The results of the process of feature extraction, refinement and categorization have been used to define and position the features as shown in Table 4.

4. Description of the Framework

The process as described in Section 3 has resulted in a framework that consists of two main parts: content coverage, which turned out to be the only application-dependent feature identified, and a description and categorization of application-independent features. First, methods to determine content

coverage are made explicit. Second, the application-independent description is presented, in which features are organized according to the 3 by 6 grid categorization that was introduced in Section 3.2.

4.1 Application-dependent Description: Content Coverage

Content coverage is one of the most important aspects of a terminological system, since physicians need to be able to completely and accurately depict the patient status or care process. Also, clinical researchers need to be able to construct patient groups at any desired level of aggregation and be ensured that all patients involved are included in these groups [26]. The content of a terminological system includes all concepts, the relationships between these concepts and the terms that describe these concepts (and relations) in natural language(s), as well as any composition rules, concept definitions and codes. Coverage of concepts and terms, which are measured in relation to the intended domain and usage, are application-dependent.

Various methods have been applied to evaluate the coverage of the concepts or terms. One example is to measure the coverage of concepts in a terminological system already in use based on the number of concepts that had to be added to the system due to under-representation in the terminological system [27].

Another approach to assess the coverage of concepts and terms is through 'concept matching' and 'term matching' [28-35]. These two measures may give different results in the situation where synonymy is supported but some synonymous terms are not included in a terminological system. For 'concept matching' and 'term matching', a representative subset of concepts and terms, respectively, is extracted from documentation in the domain of intended application. For example, if a terminological system is evaluated with regard to its use by nurses for documentation of nursing information, then the subset of concepts could be well extracted from existing nursing documentation in medical records [28, 30]. This subset of concepts or terms is then matched

with the content of the terminological system. In both term coverage and concept coverage we can distinguish between the “token” coverage, where concepts or terms are counted in accordance to their frequency of use, and the “type” coverage, in which concepts and terms contribute equally, irrespective of their frequency of use. This distinction has been made for example in [32, 35]. For nomenclatures, one can determine post-coordinated concept coverage, which takes into account both pre-coordinated (concepts as such present in the TS) and post-coordinated concepts (compositions of pre-coordinated concepts). The definitions for the various measures for content coverage are provided in Table 3.

The extent to which a concept or term can be matched with concepts in a terminological system is mostly presented as a ‘match score’ [32, 36, 37]. The coverage of the content can be represented, for example, by calculating the percentage of perfect matches, approximate matches and non-matches.

In [32] a systematic comparison is presented of the concept coverage of seven terminological systems for five “semantic domains” (i.e. “diagnoses”, “findings”, “modifiers”, “other”, and “treatments and procedures”), distinguishing “incident samples” (i.e. concept token coverage) and “unique subsets” (i.e. concept type coverage). Availability of such subsets to a broad public and reproducible methods to determine and present coverage can result in benchmarks for application-dependent assessment of terminological systems. A made-up example of the presentation of results of various content coverage measures is shown in Figure 3.

4.2 Application-independent Description

Section 3 has described the process of formulating the framework. After extracting the application-independent features from the literature, we categorized these features according to two axes. Table 4 shows the result of this categorization by type of terminological system on the horizontal axis and by elements of terminological systems

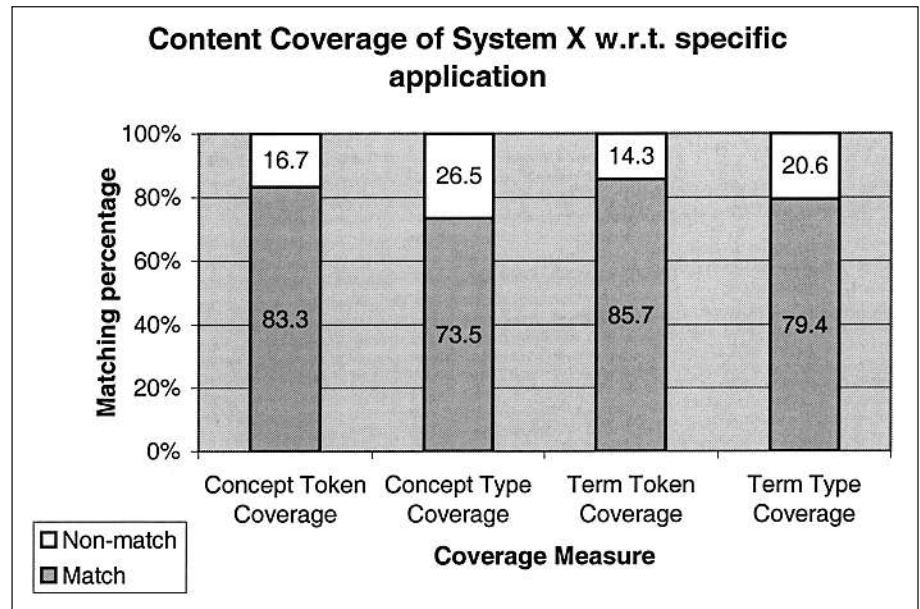


Fig. 3 Example of presentation of content coverage measurement results

and servers on the vertical axis. Features are categorized in the most applicable category. For example, the feature “number of concepts” is placed under “terminology”, as it is relevant for all (concept-oriented) terminologies.

The features in Table 4 are presented as explicit questions. Answering these questions provides a description of the application-independent characteristics of a terminological system. An example of this is described in Section 5.

5. Application of the Framework to SNOMED CT

We applied the framework as has been described above to the July 2003 UK version of SNOMED CT, used in combination with the CLUE browser 5.5. SNOMED CT was chosen as it is recommended as the foundation of a standard vocabulary in both the USA and the UK, and consequently it has recently been receiving much attention.

5.1 Application-dependent Description: Content Coverage of SNOMED CT

We have performed a provisional concept coverage study for SNOMED CT in the domain of intensive care. For this study, we used the same data set as in [38]. This data set consists of all diagnoses that formed (a part of) the in- and exclusion criteria of clinical studies that appeared in two important intensive care journals (Intensive Care Medicine and Critical Care Medicine) between January 1, 2001 and July 1, 2001. Figure 4 presents the concept and term coverage for this application. It shows that “token coverage” is higher than “type coverage”, indicating that the concepts and terms that are present in SNOMED CT are those that are used more frequently. The relatively low coverage of about 70% can be explained by the fact that many aggregations are based on highly domain-specific concepts, such as “encephalopathy with pathogenesis other than sepsis (e.g. hepatic encephalopathy)”. It is worth addressing the question of whether such concepts should be represented in a terminological system, but such a treatise is beyond the scope of this paper.

Table 4 Two-axial categorization of questions to obtain application-independent characteristics of terminological systems

	Terminology (list of terms)	Thesaurus (indexing and synonyms)	Classification (is-a relationships)	Vocabulary ((formal) definitions)	Nomenclature (composition rules)	Coding system (codes)	
Formalism	Are "concepts" and "terms" explicitly distinguished? Is length of terms restricted? Which character encoding mechanism is used? Can concepts be marked as obsolete?	Are terms indexed? Are synonyms allowed, i.e. can multiple terms have the same meanings? How is synonymy represented? Can multiple languages be represented? Are synonyms for fragments allowed? (e.g. cardiac ~ heart)	Can hierarchical relationships between concepts be defined? If yes, Which? Part-of? Is-a? Is poly-hierarchy supported? Is hierarchy restricted in depth or breadth? Can classification be inferred based on a concept's definition?	Is the meaning of concepts represented in free text? Is the meaning of concepts represented formally? If yes: how? e.g. frames, Description Logic (DL) If DL: which DL? Are relationships explicitly defined?	Is composition of concepts possible? How is this represented? Can equivalent definitions be detected automatically? Can compositions change the meaning of a concept, or do they only specify concepts in more detail?	Are codes assigned to concepts? If yes, is there code generation mechanism? Are lengths of codes restricted? Is there a meaning to these codes (e.g. mnemonic)? Do the codes limit the taxonomic placement of concepts?	Terminological system
Content	How many total concepts and terms are in the terminology? Which areas/ domains are covered?	In which way(s) are the terms indexed? In what languages are terms described?	Can properties be inherited to subordinate concepts? What is the distribution of the number of parents per concept?	Are all concepts defined/described, or only "core concepts"? e.g. diseases, but not anatomy How many concepts are: - vague? - ambiguous? - redundant? How many and which relationtypes do exist?	How many concepts can be combined or further specified?	Are all concepts coded? Are the codes proprietary or cross-mapped to another system?	
Functionality	How can terms be searched? E.g. convert code to text, keyword match, lookup phrases (incl wildcards), case insensitive, etc.	Can terms be translated from one language to another?	Can all descendants of a concept be retrieved at once?	Are multiple consistent views provided? Can properties of a concept be retrieved (e.g. definition retrieval)? Is basic inference supported, e.g. subsumption testing, instance checking?	How is a user supported in constructing composite concepts? Can refinable relations be retrieved?	Can codes be cross-mapped to codes in another coding system?	Terminology server

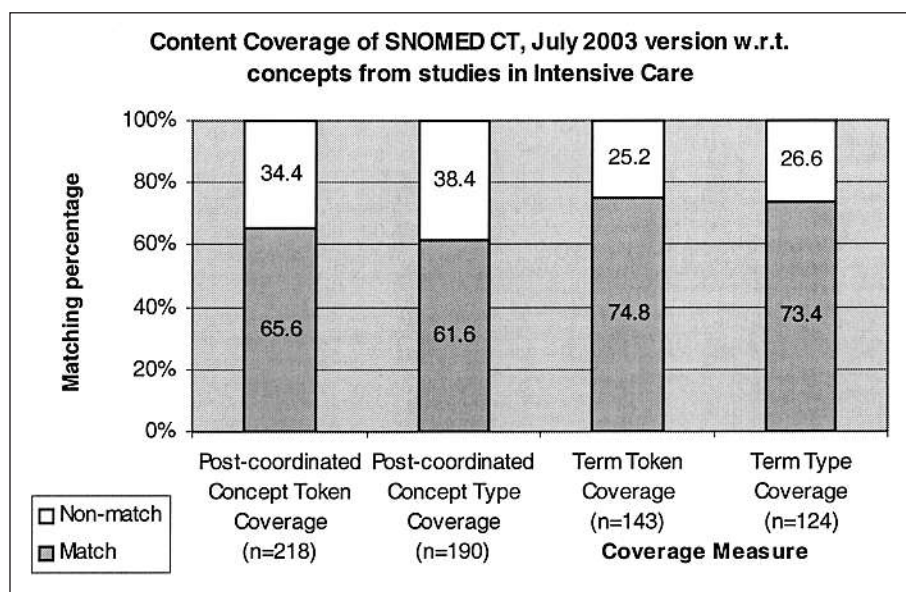


Fig. 4 Content coverage of SNOMED CT w.r.t. concepts that were retrieved from studies published in two Intensive Care journals [38]

5.2 Application-independent Description of SNOMED CT

Now we address the questions described in Table 4 for SNOMED CT. In Table 5 we summarize each question in an italic typeface, followed by a short answer. Figure 5 shows statistics of three features related to the content of the thesaurus (the number of concepts with synonymous terms), the nomenclature (the number of refinable concepts), and the classification (the number of parents per concept), respectively.

6. Discussion

The framework as described and applied in this paper aims at providing a charac-

Table 5 Two-axial categorization of questions to obtain application-independent characteristics of SNOMED CT (formalism and content) and the CLUE Browser (functionality)

	Terminology	Thesaurus	Classification	Vocabulary	Nomenclature	Coding system	
Formalism	<p><i>concepts and terms distinguished:</i> yes</p> <p><i>term length restriction:</i> none</p> <p><i>character encoding:</i> UTF-8</p> <p><i>concept obsolescence mechanism:</i> yes, concept status flag, a.o. retired, with motivations</p>	<p><i>terms indexed:</i> none</p> <p><i>supports synonymy:</i> yes, concepts can be represented by multiple terms in different languages</p> <p><i>synonym representation:</i> description status, which can be preferred, synonym, fully specified or unspecified</p> <p><i>multilingual representation:</i> yes, by means of language code</p> <p><i>description obsolescence mechanism:</i> yes, description status flag, a.o. retired, with motivations</p> <p><i>synonyms for fragments:</i> no, only for full terms</p>	<p><i>hierarchical relationships:</i> ls-a, Part-of</p> <p><i>allows polyhierarchy:</i> yes, unrestricted</p> <p><i>hierarchical depth restriction:</i> none</p> <p><i>hierarchical breadth restriction:</i> none</p> <p><i>classification inferred based on concept def:</i> yes, DL based. Included in distributed version, not supported by CLUE browser</p>	<p><i>supports free-text concept definition:</i> none</p> <p><i>supports formal concept definition:</i> yes, DL : ELH + role groups + role composition + right identity axioms</p> <p><i>atomic concepts distinguished:</i> no</p> <p><i>explicitly defined relationships:</i> yes</p>	<p><i>composition possible:</i> yes</p> <p><i>composition formalism:</i> characteristic type = 0 or 1 (0 = Defining, 1 = Qualifier), refinability = 0, 1 or 2 (0 = Not refinable; 1 = Optional: May be refined by selecting subtypes; 2 = Mandatory: Must be refined by selecting a subtype.)</p> <p><i>detection of equivalent definitions:</i> the formalism (DL) supports this, as do DL reasoners. No known tools that support postcoordination and detection of equivalent definitions</p> <p><i>compositions change meaning or only specify more detail:</i> more detail only</p>	<p><i>codes assigned:</i> yes</p> <p><i>code generation mechanism:</i> sequential number + partition identifier + check digit</p> <p><i>code length restriction:</i> 18 positions</p> <p><i>meaning of identifiers:</i> none</p> <p><i>Limitation of taxonomic placement:</i> no</p>	Terminological system
Content	<p><i>total number of concepts:</i> 352662</p> <p><i>total number of terms:</i> 939705</p> <p><i>covered areas:</i> disorders, subjective symptoms, findings, procedures, lab, radiology, anatomy, medication, chemicals, devices, care management, assessment tools</p>	<p><i>number of concepts with synonymous terms:</i> see Figure 5a (average = 2.66).</p> <p><i>languages:</i> UK English, US English, German, Spanish</p>	<p><i>properties inherited to subordinates:</i> yes</p> <p><i>number of parents per concept:</i> Figure 5c (average = 1.30).</p>	<p><i>approx number of vague concepts:</i> 13151 (3.7%) (based on preferred terms containing " or " or "/or")</p> <p><i>approx number of ambiguous concepts:</i> n/a</p> <p><i>approx number of redundant concepts:</i> n/a</p> <p><i>relationships used:</i> 993 attributes are defined, 42 are used.</p>	<p><i>number of refinable concepts:</i> 153080 (43%)</p> <p>distribution of refinable relations per concept see Figure 5b</p>	<p><i>All concepts coded:</i> yes</p> <p><i>crossmappings:</i> CTV-3, CDT-2, HHCC, ICD-9-CM, ICD-10, ICD-O, LOINC, NIC, NANDA, PNDS, OMAHA, OPCS-4</p>	
Functionality	<p><i>convert code to text:</i> only for SNOMED concept- and description lds</p> <p><i>lookup phrases for a string:</i> yes</p> <p><i>lookup phrases matching a string (with wildcards); inexact match:</i> yes, works on parts of words</p> <p><i>code refinement:</i> no</p> <p><i>keyword matching:</i> no</p> <p><i>case-insensitive:</i></p>	<p><i>translation to other languages:</i> no, one language-specific version is used</p>	<p><i>retrieve descendants:</i> yes</p>	<p><i>providing multiple consistent views:</i> yes, subsets can be defined by means of reference to concepts to be included or excluded (with or without their subsumees)</p> <p><i>retrieve definitions:</i> yes</p> <p><i>subsumption testing:</i> no</p> <p><i>instance checking:</i> no</p> <p><i>detection of equivalent definitions:</i> no</p> <p><i>query for concepts matching structural criteria:</i> no</p>	<p><i>support in concept composition:</i> no</p> <p><i>retrieve refinable relations:</i> yes</p>	<p><i>cross coding:</i> no</p>	Terminology service

terization of terminological systems. This characterization needs to strike a balance between conciseness and completeness. A complete characterization is impractical if not impossible, as there may always remain new features that can be defined. Hence, the features defined in this paper provide a good

starting point, not a definitive collection. The process we followed to the identified features makes it fair to assume that these features are indeed important ones.

The distinction between formalism, domain knowledge and functionality helps to identify the strengths and weaknesses of

systems, and the possibility to overcome the weaknesses. Generally, shortcomings in the content can be solved relatively easily, whereas shortcomings in the formalism are harder to overcome. Likewise, if a terminology server lacks functionality, this can only be implemented if the formalism underlying

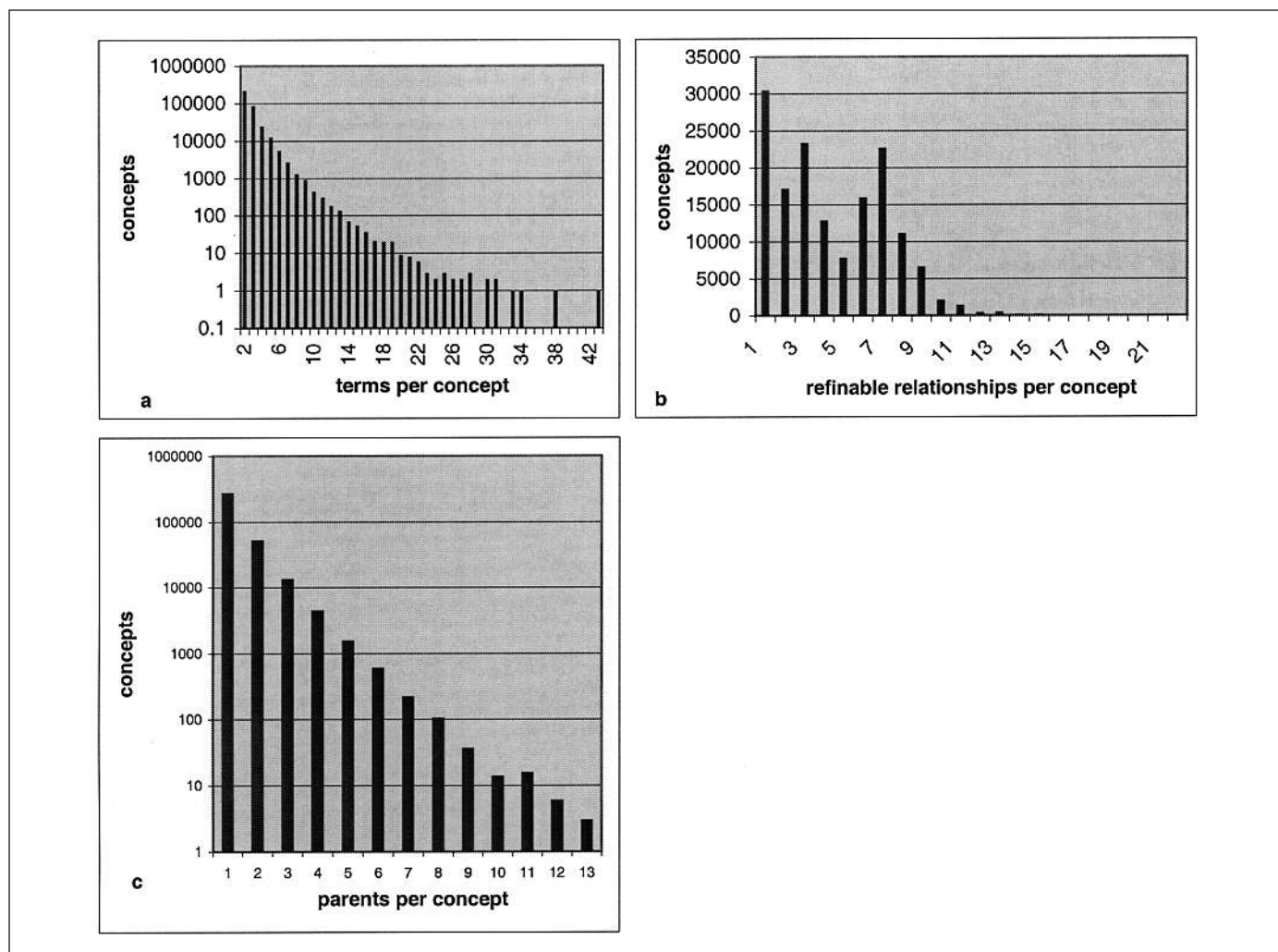


Fig. 5 Use of terms and compositionality in SNOMED CT. Figure a shows a histogram of the number of synonyms (on a log scale), Figure b shows the histogram of the number of refinable relationships (the composition mechanism in SNOMED CT), Figure c shows the histograms of the number of parents.

a terminological system provides support for such functionality. For example, to provide word normalization, the formalism should allow for the representation of normal forms and inflections of terms.

In this section we will further discuss the limitations and possible drawbacks of this framework by looking at various application tasks of the framework: comparison between terminological systems, fulfilment of requirements, and development of a terminological system. We will furthermore relate the framework to the literature, and look at the possibilities for and benefit of sharing experiment results using this framework.

6.1 Using the Framework for Comparing Terminological Systems

The utility of the framework presented in this paper increases if researchers and developers of terminological systems would address the questions described in Table 4 and make their answers publicly available. The availability of a structured characterization of various terminological systems will support their comparability but some problems will still remain. The first problem is that, although the feature values are described, the interpretation of their implication may be difficult. For example the representation formalisms of terminologi-

cal systems can be described as different description logics, but it may be hard to interpret what the (practical) consequences are of these different formalisms. To further enhance comparability, not only the features should be explicitly specified, but also their allowed values i.e. the possible feature values, for example “DL, frames, other” for the feature “Formalism used”. Currently, no categories for feature values are presented; instead they are specified in free-text. Secondly, some features in the categorization are hard to measure, such as the number of vague, ambiguous or redundant concepts. Thirdly, measurement of the application-dependent feature “content coverage” remains labor-intensive, and should be performed

for each domain and application, as existing subsets may not always be representative for intended new usage. It is important that these subsets are made publicly available so that similar subsets can be used to evaluate content coverage of different terminological systems.

6.2 Using the Framework for Requirements Fulfilment

Figure 2 depicts how features and their values can be used in the process of selecting a terminological system for a specific application. These feature values need to be matched with requirements that the application poses on the system. This is not necessarily a straightforward process. One needs to determine the domain of interest (e.g. intensive care) and the application (e.g. recording the reason for admission). Three types of application of a terminological system can be distinguished according to [39]: 1) entering and presenting data about patients, 2) sharing and integrating information, and 3) querying and retrieving information. The NHS Information Authority describes nine scenarios that were mentioned in Section 1, which we will place into these three categories. As a rule of thumb, the following requirements hold for these applications:

- *Entering and presenting data about patients* (NHS: Documentation in the EPR/EHR, Decision support): a terminological system should capture all concepts and terms that are applicable. Either these concepts must be present in the system, or it must be possible to compose them using pre-existing concepts, a process also called post-coordination. The latter poses demands on the nomenclature-related features of a terminological system. Properly handling terms entered by clinicians involves terminology- and thesaurus-related features of the system.
- *Sharing and integrating information* (NHS: Clinical audit, Reporting, Summaries, Decision support): semantics of the data should be well understood by all people involved and by other involved software applications, such as decision support systems. Hence, definitions are

important (vocabulary-related features). In addition, it may be desirable to present information at varying levels of detail, especially when information is exchanged between various specialties, for example between a cardiac surgeon and a general practitioner. This poses demands on the classification-related features.

- *Querying and retrieving* (Administrative and management information, Epidemiology, Billing, and Resource management): this requires the ability to aggregate data (classification-related features). It needs to be determined how concepts need to be aggregated, e.g. whether there are predefined axes (sometimes referred to as chapters), or whether it should be possible for users to freely combine concept-criteria. The latter case requires concepts to be explicitly characterized by properties and relations to other concepts (i.e. vocabulary-related properties).

These descriptions of the types of applications indicate which terminological system types (e.g. vocabulary, nomenclature) are the most relevant. Explicit categorization by the terminological system type (represented by the columns of Table 4) supports focusing on those features that are essential for fulfilling the requirements of the terminological system. A more detailed requirement analysis is outside the scope of this paper.

6.3 Using the Framework for Development of a Terminological System

Many contemporary systems have evolved over time by adding contents, broadening the scope, and aiming at supporting a wider range of applications. These systems often provide workarounds for limitations of previous versions, without really solving the underlying problem. For example, ICD10 has introduced the “dagger/asterisk” mechanism to enable dual coding, which is a partial solution for creating a poly-hierarchy, which was not supported in earlier versions. A clear understanding of the initial and fu-

ture use of a terminological system, together with the requirements that emerge from this usage, can help determine the necessary features and feature values that the terminological system should possess. We believe that the framework described in this paper will also contribute to development of terminological systems from scratch, as the framework encourages developers to determine which features are important for their specific application.

6.4 Relation between Framework and Literature

In this section we will relate our work to the literature on which we based our framework, although we do not claim that this bibliography provides a complete overview.

The desiderata specified in [14] focus on application-independent and application-dependent characteristics. These mainly involve formalism-related issues (concept orientation, concept permanence, nonsemantic identifiers, polyhierarchy, formal definitions, rejection of “Not Elsewhere Classified”, multiple granularities, multiple consistent views, representation of context). “Content” in itself is defined as the most important characteristic of terminological systems. However, methods to evaluate this application-dependent characteristic, as summarized in section 4.2, are not provided in [14]. “Recognition of redundancy” relates both to functionality and formalism for detection of equivalent (post-coordinated) concept definitions. “Graceful evolution” is not within the scope of this paper since it involves a formalized organization for keeping track of changes between versions of terminological systems, and in our framework we do not take maintenance and versioning issues into account.

The Quality Indicators from [15] cover the application-independent characteristics as mentioned in [14] but add more detail to these characteristics, and provide some additional characteristics, such as: clearly stated “purpose and scope” of terminological systems, and functionality for “normalization of content and semantics”. It is furthermore stated that composition of con-

cepts must be possible, i.e. that a terminological system is a nomenclature. A notable issue mentioned in [15] is the need for specification of some application-specific requirements, such as: persistence and extent of (primary) use, and the degree of automatic inference intended. As described in Figure 2, these application-specific requirements should be mapped to the feature values of a terminological system.

LQS [18], being a specification for implementation of a terminology server, can also be used as a structured reference for required functionality. In this way it can play an important role for the application-independent description of functionality of existing terminology servers, and guide development of new terminology servers. The functions described in LQS outnumber the functionality defined in our framework, as we defined functionality at the level of use cases rather than function calls in an Application Programming Interface (API). This restriction is motivated by the need to keep the framework at a conveniently high level of granularity. The development of new terminology server interfaces such as HL7 Common Terminology Services (CTS)^d may be a next step towards providing means for the characterization and implementation of terminology servers, and provide a valuable addition to our framework.

The National Committee on Vital and Health Statistics (NCVHS) Questionnaire [17] is the first effort known to the authors that delivers a structured, application-independent description of a variety of terminological systems. If the results become available electronically it will provide a valuable source of information for comparison, evaluation or development of terminological systems. The main added value of our framework compared to [17] is the categorization of features, and the provision of explicit measures and methods for content coverage.

By making characteristics explicit and by striving to make them objectively measurable, the barriers to evaluation of terminological systems as described by Hales

[16] can be overcome. The first three barriers “1) evaluations are application dependent, 2) assessment is empirical instead of independent, 3) dichotomous measures of characteristics (presence or absence) are more in use than continuous measures” have been overcome by using our general approach presented in Figure 2. By using a structured categorization of characteristics (Table 4) in the first phase of this approach we paid attention to the remaining barriers “4) poor definition of characteristics, 5) large number of characteristics, 6) different significance of characteristics, and 7) interdependence of characteristics”.

6.5 Reuse of Results – Possibilities and Benefits

The framework presented in this paper provides a template for organizing features of terminological systems. This is of importance for increasing the understanding of terminological systems. However, its additional value is in the application of the framework, and in sharing the results. The usefulness of the framework will increase with the number of terminological systems and applications it is applied to. If descriptions are shared and detailed methods are made available, comparison, evaluation and development of TSs will become less complex than it currently is. As making an application-independent description is a one-time effort for each (version of a) terminological system, and requirements specification is a one-time effort for each application, this framework can contribute to a reduction of the effort to be put in comparison and evaluation, assuming that researchers share their results. This sharing of results requires a repository in which the characterizations of individual (versions of) terminological systems are held. Providing an open (e.g. web-based) environment will increase the utility of this framework. Firstly, such an environment can provide more information about systems, their feature values, and the methods to determine these feature values. Secondly, if results are shared, the relative cost of comparison of terminological systems will decrease. Finally, performing such comparisons may

provide insight into the most prominent possibilities for improvement.

The realization of such an infrastructure comprises further work we are planning to undertake.

7. Conclusion

The framework described in this paper aims to describe the essential characteristics of the terminological systems. This enhances the understanding of these systems, which is necessary for comparison, application, and development of terminological systems. Since most characteristics are application-independent, their description can be reused in different applications. The proposed categorization as described in Table 4 supports the explicit “once-only” description of the application-independent characteristics of a terminological system. Thereby, this framework aims to reduce the efforts for determining which terminological system is applicable for a certain clinical setting. This framework may also help terminological system developers to determine in what way their system can be improved to serve more or broader needs.

We have combined the two axes “Elements of terminological systems and servers” and “Types of terminological systems”, resulting in a 3 by 6 grid of application-independent characteristics as presented in Table 4. In this grid we have specified questions as examples of the characteristics of terminological systems. These questions are examples of how to specify the application-independent description of terminological systems, the first step in the evaluation process. We plan to realize a website through which the set of questions is made available. In this way, characterizations of individual terminological systems as well as future additions to the current set of questions can be made publicly accessible.

The process described in this paper reveals the prominent features of terminological systems. The framework provides a structured categorization of features that constitute characterization of a terminological system. Beyond understanding termino-

^d <http://informatics.mayo.edu/index.php?page=11>
(last visited February 14, 2005)

logical systems, the main application tasks that this framework supports are comparison of terminological systems, fulfilment of requirements, and development of a terminological system.

The application of the framework to SNOMED CT and the CLUE Browser demonstrates the applicability of the framework. Further application of the framework to a variety of terminological systems and servers will help to increase the understanding of their individual merits, and eases the process of determining the applicability of a terminological system for a specific application and domain.

Acknowledgments

This work is supported by the Netherlands Organization for Scientific Research (NWO) program "Information & Communication Technology in Healthcare" (ICZ) for the project entitled "Terminology and Semantics: Making semantics explicit", number 014-18-014.

References

- de Keizer NF, Abu-Hanna A, Zwetsloot-Schonk JH. Understanding terminological systems. I: Terminology and typology. *Methods Inf Med* 2000; 39 (1): 16-21.
- Rossi Mori A, Consorti F, Galeazzi E. Standards to support development of terminological systems for healthcare telematics. *Methods Inf Med* 1998; 37 (4-5): 551-63.
- International Statistical Classification of Diseases and Related Health Problems, 1989 Revision. Geneva: World Health Organization; 1992.
- Rogers FB. Medical subject headings. *Bulletin of the Medical Library Association* 1963; 51: 114-6.
- Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug safety: an international journal of medical toxicology and drug experience* 1999; 20 (2): 109-17.
- McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical Chemistry* 2003; 49 (4): 624-33.
- Coté RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, editors. SNOMED International: the systematized nomenclature of human and veterinary medicine. Vols I-IV. Northfield, IL: College of American Pathologists; 1993.
- Spackman K. SNOMED RT and SNOMED CT. Promise of an international clinical terminology. *MD Computing* 2000; 17 (6): 29.
- Rector AL, Nowlan WA. The GALEN project. *Computer Methods and Programs in Biomedicine* 1994; 45 (1-2): 75-8.
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* 2004; 32 (Database issue): D258-61.
- Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics* 2003; 36 (6): 478-500.
- Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. *Bulletin of the Medical Library Association* 1993; 81 (2): 217-22.
- Ingeneff J, Giere W. Concept-oriented standardization and statistics-oriented classification: continuing the classification versus nomenclature controversy. *Methods Inf Med* 1998; 37 (4-5): 527-39.
- Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 1998; 37 (4-5): 394-403.
- ISO/TC215 WG 3. Standard Specification for Quality Indicators for Controlled Health Vocabularies; 2000 July. Report No.: TS17117.
- Hales JW, Schoeffler KM. Barriers to Evaluation of Clinical Vocabularies. In: Cesnik B, McCray AT, Scherrer JR, editors. *Medinfo1998: Proceedings of the 9th World Congress on Medical Informatics*; 1998; Seoul, Korea. Amsterdam, The Netherlands: IOS Press; 1998. pp 680-4.
- NCVHS SSS. Summary and Analysis of Terminology Questionnaires Submitted by Developers of Candidate Terminologies for PMRI Standards. draft; 2003 April 17.
- OMG. Lexicon Query Service Specification: Object Management Group; 2000 July. Report No.: 00-06-31.pdf.
- Smith B. Beyond Concepts: Ontology as Reality Representation. In: Varzi A, Vieu L, editors. *Proceedings of FOIS 2004. International Conference on Formal Ontology and Information Systems*; 2004; Turin: Amsterdam, The Netherlands: IOS Press; 2004. pp 73-84.
- Bodenreider O, Smith B, Burgun A. The Ontology-Epistemology Divide: A Case Study in Medical Terminology. In: Varzi A, Vieu L, editors. *Proceedings of FOIS 2004. International Conference on Formal Ontology and Information Systems*; 2004; Turin. Amsterdam, The Netherlands: IOS Press; 2004. pp 185-95.
- Supekar K. A Peer-review Approach for Ontology Evaluation. In: Noy NF, editor. *8th International Protégé Conference*; July 18-21, 2005. Madrid, Spain; 2005. pp 77-9.
- Supekar K, Patel C, Lee Y. Characterizing quality of knowledge on semantic web. In: Barr V, Markov Z, editors. *Proceedings of the Seventeenth International FLAIRS Conference*; May 2004; Miami, FL: AAAI Press; 2004. pp 220-8.
- Cornet R. Towards Structured Requirements for Terminological Systems and Servers. In: Patel V, Rogers R, Haux R, editors. *Medinfo 2001: Proceedings of the 10th World Congress on Medical Informatics*; 2001; London, UK: IOS Press, Amsterdam, The Netherlands; 2001. p 295.
- CEN/TC251. Medical Informatics – Categorical structures of systems of concepts – Model for representation of semantics. European Prastandard. Brussels, Belgium: European Committee for Standardization; October 1997. Report No.: ENV 12264.
- Guarino N. Formal Ontology in Information Systems. In: Guarino N, editor. *Proceedings of FOIS'98*; June 1998; Trento, Italy: IOS Press, Amsterdam, The Netherlands; 1998. pp 3-15.
- Arts DGT, Cornet R, de Jonge E, de Keizer NF. Methods for Evaluation of Medical Terminological Systems. *Methods Inf Med* 2005; 44: 616-25.
- Bodenreider O, Burgun A, Botti G, Fieschi M, Le Beux P, Kohler F. Evaluation of the Unified Medical Language System as a medical knowledge source. *Journal of the American Medical Informatics Association* 1998; 5 (1): 76-87.
- Bakken Henry S, Holzemer WL, Reilly CA, Campbell KE. Terms used by nurses to describe patient problems: can SNOMED III represent nursing concepts in the patient record? *Journal of the American Medical Informatics Association* 1994; 1 (1): 61-74.
- Bodenreider O, Mitchell JA, McCray AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. In: Kohane IS, editor. *Proceedings of the 2002 AMIA Annual Symposium*; 2002; San Antonio, TX, USA. Philadelphia, PA: Hanley and Belfus Inc., USA; 2002. pp 61-5.
- Bowles KH. Application of the Omaha System in acute care. *Research in Nursing & Health* 2000; 23 (2): 93-105.
- Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity. *Journal of the American Medical Informatics Association* 1997; 4 (3): 238-51.
- Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR, For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. The content coverage of clinical classifications. *Journal of the American Medical Informatics Association* 1996; 3 (3): 224-33.
- Cimino JJ, Patel VL, Kushniruk AW. Studying the human-computer-terminology interface. *Journal of the American Medical Informatics Association* 2001; 8 (2): 163-73.
- Hardiker NR, Rector AL. Structural validation of nursing terminologies. *Journal of the American Medical Informatics Association* 2001; 8 (3): 212-21.
- Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. *Journal of the American Medical Informatics Association* 1997; 4 (6): 484-500.
- Warnekar P, Carter J. HIV Terms Coverage by a Commercial Nomenclature. In: Musen MA, Informatics; 2001; London, UK: IOS Press, Amsterdam, The Netherlands; 2001. p 295.

- Friedman CP, Teich JM, editors. Proceedings of the 2003 AMIA Annual Symposium; 2003; Washington, DC, USA. Philadelphia, PA, USA: Hanley and Belfus Inc.; 2003. p 1046.
37. Wasserman H, Wang J. An Applied Evaluation of SNOMED CT as a Clinical Vocabulary for the Computerized Diagnosis and Problem List. In: Musen MA, Friedman CP, Teich JM, editors. Proceedings of the 2003 AMIA Annual Symposium; 2003; Washington, DC, USA. Philadelphia, PA, USA: Hanley and Belfus Inc.; 2003. pp 699-703.
38. Arts D, De Keizer N, De Jonge E, Cornet R. Comparison of methods for evaluation of a medical terminological system. In: Fieschi M, Coiera E, Li J, editors. Proceedings from Medinfo 2004, San Francisco, CA, USA. Amsterdam, The Netherlands: IOS Press; 2004. pp 467-71.
39. Rector AL. Clinical terminology: why is it so hard? *Methods Inf Med* 1999; 38(4-5): 239-52.

Correspondence to:

Ronald Cornet
Department of Medical Informatics, J1b-114
Academic Medical Center, Universiteit van Amsterdam
P.O. Box 22700
1100 DE Amsterdam
The Netherlands
E-mail: r.cornet@amc.uva.nl