Methods for auditing medical terminological systems
Cornet, R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Auditing DL-based Medical Terminological Systems by Detecting Equivalent Concept Definitions

Ronald Cornet, Ameen Abu-Hanna

Under submission
Abstract

Objective To devise and evaluate a method for auditing Medical Terminological Systems (TSs) based on detecting concepts with equivalent definitions. Our method addresses two important problems: redundancy, where the same concept is represented more than once, and underspecification, where different concepts have the same representation and hence appear indistinguishable from each other.

Design The auditing method is applicable for TSs that are or can be represented in a description logic (DL). The method comprises the assumption that concept definitions are non-primitive (i.e. they are regarded as providing necessary and sufficient conditions). Whereas this assumption may not be correct for many definitions, it does serve the purpose of detecting sets of logically equivalent concepts by a DL reasoner. These sets indicate concepts that are defined more than once, and concepts that are indistinguishable from each other by their properties, as they are underspecified. Analysis of these sets provides insight into the representation quality of concepts and provides hints at improving the TS.

Measurements In our case study the method is applied to the DICE TS, a comprehensive TS in Intensive Care. It comprises about 2500 concepts and 40 properties and relations.

Results In DICE we found four concepts that were defined twice. Furthermore, 100 sets were found containing more than 300 underspecified concepts. The sizes of these sets ranged from two to thirteen. Analysis revealed that many concepts can be more completely defined, either by adding existing relations, or by the introduction of new relations into the terminological system.

Conclusion The method proved both usable and valuable for auditing TSs. DL reasoning is fully automated and all equivalent concept definitions are systematically found. The resulting sets of equivalent concepts clearly point out which concept definitions are to be reviewed, as they contain duplicate definitions of a concept, and (inherently or unnecessarily) underspecified concepts.

6.1 Introduction

Medical terminological systems (TSs) represent knowledge by means of concepts, relationships and terms in a medical domain. For example, in a TS a concept may be defined as “inflammation of the membranes of the brain or spinal cord”, and described by the synonymous terms “cerebrospinal meningitis” and “meningitis”. TSs provide an invaluable source of (structured) medical knowledge, and have developed from single-purpose systems to systems serving a range of purposes, varying from recording patient information to providing decision support, and supporting epidemiological research and resource management. In order to be able to address this range of purposes, medical TSs have grown in size and complexity [1]. They evolved from simple taxonomies to semantic networks with (informal and formal) concept definition capability. Es-
Especially during the last decade, formal concept definition has gained attention. Such definitions are commonly represented using frames and description logics (DLs). Examples of frame-based TSs are the Foundational Model of Anatomy (FMA) [2] and the Gene Ontology (GO) [3]. Examples of DL-based TSs are SNOMED-CT [4], the National Drug File Reference Terminology (NDF-RT) [5], and the NCI Thesaurus [6].

An important requirement for TSs is that the represented knowledge should be of good quality especially in terms of its internal consistency and faithfulness to reality. A formal representation provides explicit semantics of the represented knowledge, thus facilitating the determination of the consistency of this knowledge, and helping in checking its faithfulness. The process of quality assessment is called auditing. In the next section we will discuss a number of auditing approaches that have been designed and applied to ontologies in the field of medicine. In these approaches computational methods are used to focus the attention of a modeler to suspicious definitions, after which a modeler analyzes these definitions and eventually modifies the represented knowledge.

We focus in this paper on the automatic discovery of equivalently defined concepts, which might correspond to duplicate concept definitions or underspecified concepts. Duplicate concept definitions are undesirable [7], and should not occur in a TS, because they may hamper querying a TS. For example, if a TS would contain “myocardial infarction” and “heart attack” as separate concepts, and it is used to query for patients with heart attack, patients registered as having a myocardial infarction will not be returned. Underspecified concepts that appear indistinguishable from each other can be analyzed to determine whether it is possible to sharpen their definitions.

Description logics form our representation and reasoning machinery. Our goal is to explore the possibilities of deploying DLs in the audit of concept definitions in a TS. The DL family was chosen because its formal representation allows performing tractable automated reasoning on the represented knowledge. The prominent reasoning services are satisfiability testing (i.e. the logical consistency of represented knowledge) and subsumption (i.e. classification of a concept based on its properties). For detecting equivalence we will use the DL reasoning service of subsumption testing. Concepts that are rendered equivalent by a DL reasoner can then be further analyzed. If the logical equivalence is due to a duplicate definition of a concept, such a definition can be removed. If equivalence is caused by underspecification, it needs to be determined whether a more elaborate specification of a concept can be given.

We provide an overview of auditing approaches and an introduction to DL in Section 6.2 and then explain our method in detail in Section 6.3. The results of the application of our method in a case study in the intensive care domain are discussed in Section 6.4. Section 6.5 summarizes the general results of the application of our method. Finally, conclusions are drawn in Section 6.6.
6.2 Background

Modeling large knowledge bases and evaluating their contents are complicated processes. The need arises for systematic, reproducible methods to support these processes. Modeling and evaluating TSs concern various aspects, ranging from ontological decisions to determining the comprehensiveness of the medical contents of a TS. Ideally, a knowledge base should satisfy four requirements [8]: (1) it should have the necessary knowledge (completeness), (2) the knowledge should be faithful to the real world (correctness), (3) the knowledge should not be self-contradictory (consistency), and (4) the system should have efficient algorithms to perform the inferences needed for the application (competence). Auditing is the process of assessing the fulfillment of (one or more of) these requirements. After discussing a number of auditing approaches, we will introduce description logics and then address the use of description logics in medicine.

6.2.1 Auditing approaches

During the last decade, various techniques have been applied for auditing medical TSs. In [9], so-called “semantic methods” are applied for the detection of: ambiguity; redundancy of concept pairs; inconsistency of parent-child relationships; and lack of semantic links. These methods make use of synonyms and of semantic types that are assigned to concepts. In [10], methods for finding missed synonymy are described, based on lexical techniques and the use of synonymous words and phrases. In [11], a technique is presented to audit concept categorizations (i.e. the assignment of one or more semantic types to a concept), based on expert review of intersections of semantic types. In [12], “semantic refinement” is presented, which helps detection of ambiguity, non-uniform classification, classification errors, omissions, redundant classification and missed synonymy. This method also makes use of semantic types. In [13] the use of Protégé Axiom Language (PAL) queries in Protégé is described for the detection of redundantly defined is_a relations. The same environment is used in [14] for the purpose of detecting constraint violations, such as concepts that have multiple preferred terms in a language, whereas exactly one preferred term is required. In [15], two algorithms (lexical comparison and classification) are combined to detect (among others) improper assignment of relationships, redundant concepts, and omission of relationships.

In these approaches the modeler eventually does a manual interpretation of parts of the represented knowledge, where the computational methods help focus attention on possible errors or flaws. In our approach the interpretational burden is further shifted towards the method itself by means of automated reasoning. Potential duplicates and underspecification are automatically detected and the modeler has to decide whether or not they constitute actual duplicate definitions or underspecified concepts, and act accordingly.
6.2.2 Description logics

Description logics (DLs) provide fragments of first order logic for formal definition of concepts. These definitions can specify either only necessary conditions or both necessary and sufficient conditions. Definitions with only necessary conditions are indicated by the notion “primitive definitions” [16, Chpt. 9] (by others also called “specialization” or “partial class”). Primitive definitions are indicated by the “subsumed by” symbol: ⊑. Definitions with both necessary and sufficient conditions are referred to as “non-primitive definitions” (by others also called “definition” or “complete class”) and indicated by the “equivalence” symbol: ≡. As an example of a non-primitive definition, axiom 1 in Figure 6.1 states that every Inflammatory Disease is necessarily and sufficiently a disease in which some inflammation is involved. This implies that every disease that involves an inflammation can be inferred to be an inflammatory disease. Axiom 5 in Figure 6.1 shows an example of a primitive definition: a ViralHepatitis1 is an inflammatory disease that is located in the liver (and maybe of other body parts). However, it can not be inferred that every inflammatory disease of the liver is a ViralHepatitis1 (as it might have a non-viral cause, for example a bacterium).

Each DL is characterized by the constructors it allows for. Examples of concept constructors are AND (⊔), OR (⊔), NOT (¬), SOME (∃), ALL (∀), AT-LEAST (≥).

The formal, set-theoretic semantics of DLs provide statements with an unequivocal meaning, although these statements are restricted by the expressiveness of the underlying DL. The foremost reasoning tasks with DLs are satisfiability testing and subsumption (classification). Satisfiability testing is checking whether a concept expression does not necessarily denote the empty concept [16]. Subsumption testing amounts to checking whether one concept is more general than another. Subsumption can be inferred by virtue of non-primitive definitions only, as these specify both necessary and sufficient conditions, as explained above. The computational complexity increases with the expressiveness of a DL. Generally, reasoning with inexpressive DLs is tractable, whereas reasoning with very expressive DLs can become intractable, and even undecidable (meaning that reasoning may require an infinite amount of time or memory).

6.2.3 Description logics in medicine

The domain of medicine consists of many natural kinds, for which no necessary and sufficient conditions exist [17]. These natural kinds are recognized rather than inferred. As these natural kinds can only be defined in a primitive manner [18], much of the inferential potential is lost, as no concept can be inferred to be subsumed by a concept with a primitive definition.

Apart from natural kinds that result in primitive concept definitions, there may be other reasons why concepts are defined as primitive, for example the expressive power of DL can be too limited to express the necessary and sufficient conditions.
6.2. Background

1. InflammatoryDisease $\equiv$ Disease $\sqcap$ $\exists$ involves Inflammation
2. LiverDisease $\equiv$ Disease $\sqcap$ $\exists$ location Liver
3. Hepatitis1 $\equiv$ InflammatoryDisease $\sqcap$ $\exists$ location Liver
4. Hepatitis2 $\equiv$ LiverDisease $\sqcap$ $\exists$ involves Inflammation
5. ViralHepatitis1 $\sqsubseteq$ InflammatoryDisease $\sqcap$ $\exists$ location Liver
6. ViralHepatitis2 $\sqsubseteq$ LiverDisease $\sqcap$ $\exists$ involves Inflammation
7. ViralHepatitisA $\equiv$ InflammatoryDisease $\sqcap$ LiverDisease $\sqcap$ $\exists$ cause HepatitisAVirus

Fig. 6.1: Concept definitions exemplifying the use of primitive and non-primitive definitions. Primitive definitions are indicated by use of the $\sqsubseteq$ symbol, non-primitive definitions by $\equiv$. The $\sqcap$ symbol states logical conjunction (i.e. “AND”), the $\exists$ symbol (“SOME”) preceding a role specifies existence of a relation with the role-filler.

In contemporary terminological systems the majority of concept definitions are primitive. In both SNOMED CT (July 2005 edition) and the NCI Thesaurus (release 05.05d) non-primitive definitions amount to only 11% of the total number of concepts.

The large number of primitive concepts reduces the possibility of finding any concepts that are defined more than once. Before we present a method to overcome this, we will first give an example to demonstrate this.

Figures 6.1 and 6.2 provide an example of the role that non-primitive and primitive definitions play in modeling of terminological systems. Non-primitive definitions facilitate the inference of classification (subsumption) of concepts. These definitions also make it possible to detect equivalent definitions of concepts (by means of detecting mutual subsumption). For example, in Figure 6.1 the first definition states that an InflammatoryDisease is a Disease that involves an Inflammation. As this is a non-primitive definition, these are necessary and sufficient conditions. Hence, any Disease that involves an Inflammation is inferred to be an InflammatoryDisease. Likewise for definition 2, a disease that is located in the liver is a LiverDisease. Hepatitis1 and Hepatitis2 (definitions 3 and 4) can be inferred to be equivalent by a Description Logic Reasoner, such as Racer$^1$, Fact++$^2$ or Pellet$^3$. It is then to the modeler to decide whether Hepatitis1 and Hepatitis2 are actually duplicate definitions of one concept, or different concepts which are equivalently defined. However, when definitions are primitive, as those of ViralHepatitis1 and ViralHepatitis2 (5 and 6), equivalence will not be inferred, and a possible duplicate definition remains undetected. Another pitfall of primitive definitions is that they can lead to missed classification. For example, given definition 7, ViralHepatitisA would be correctly classified as a child of Hepatitis1 and Hepatitis2, but not as a child of ViralHepatitis1 or ViralHepatitis2, although it should have been.

---

$^1$ http://www.racer-systems.com/
$^2$ http://owl.man.ac.uk/factplusplus/
6.3 Method

Although it is inevitable to have many primitive definitions in a Medical TS, the examples in Section 6.2 demonstrate the potential of exploiting the inferential powers of DL reasoners in the modeling process by stating the non-primitivity of all relevant concept definitions.

Our method for determining equivalent definitions comprises of the following steps, which we describe below: determine concepts of interest, exclude poorly defined concepts, assume non-primitive definitions, infer equivalence, interpret the results.

**Determining concepts of interest** The first step is to determine which concept category is to be audited. Generally, medical TSs can be regarded as consisting of various (more or less explicitly distinguishable) modules [20, 21]. For example SNOMED CT specifies not only concepts in the category “disease”, but, among others, also the categories “body structure”, “finding”, “organism”, “specimen”, and “substance”. These modules are used for the definition of disease concepts, as is also shown in the examples in Figures 6.1 and 6.3.

The need to determine which category will be investigated is driven by the fact that equivalence of concepts can propagate, leading to equivalence of other concepts. An example of this situation is presented in Figure 6.3. If we would apply our method to diseases as well as microorganisms, Virus and Bacterium would be rendered equivalent, as their definitions are the same. This in turn would lead to equivalence of ViralPneumonia and BacterialPneumonia, due to their reference to respectively Virus and Bacterium. Hence, one either focuses on microorganisms, which will point out equivalence of Virus and Bacterium, or on...
diseases, in which case ViralPneumonia and BacterialPneumonia will correctly be regarded non-equivalent.

**Exclusion of poorly defined concepts** The next step is to find all concepts that are subsumed by one concept and do not specify any difference with their subsumer. In Figure 6.3, Pneumococcal Pneumonia and Staphylococcal Pneumonia are examples of such concepts. For these concepts, changing their definitions to non-primitive definitions will provide a trivial equivalence of the concept and their subsumer. As concepts of this form are easily recognizable, they can be studied separately. An analysis on SNOMED CT [22] showed that in some modules (e.g. “organism” and “substance”) there was not a single concept that specifies any difference with its direct subsumer, whereas in other modules the proportion of concepts that explicitly specify differences with their subsumer is up to 86% (e.g. “specimen”).

**Assuming non-primitive definitions** We can now redefine all other concepts (i.e. those that are subsumed by more than one concept or show differences with their subsumer(s)) as non-primitive. In Figure 6.3, the definitions for Pneumonia and Pulmonary Edema are changed from primitive definitions to non-primitive definitions. Viral Pneumonia and Bacterial Pneumonia, which already had a non-primitive definition, remain unchanged.

**Inference of equivalence** When the TS has been altered according to the steps mentioned above, it can be classified with a DL reasoner. This classification will result in sets of equivalent concepts. These sets can then be further analyzed. Classification of the example system from Figure 6.3 will render Pneumonia and Pulmonary Edema equivalent.

**Interpretation of equivalence** At this stage it is up to the modeler to analyze the equivalences. This analysis will provide two types of outcomes. First, it will reveal concepts that have duplicate definitions, which were previously undetected due to the fact that definitions were primitive, analogous to the example of ViralHepatitis1 and ViralHepatitis2 in Figure 6.1. Second, it will reveal concepts that are different (as in the above example of Pneumonia and Pulmonary Edema), but for which the distinction between them is not represented. In the latter case, which we refer to as underspecification, the TS can potentially be enriched by making explicit the implicit knowledge that distinguishes one concept from another. When this distinction cannot be made explicit, it is due to the lack of characteristic features of the concept (i.e. it is a natural kind), or due to limitations of the DL used.
6. DL-based Auditing: Equivalence

Fig. 6.3: Example of application of the method to concepts in a “disease” category. On the left-hand side, the original representation is shown. The right-hand side shows the resulting representation, where the defining characteristics, which remain unchanged, are indicated by “...”. In the middle, the steps that involve this result are given.

6.4 Case Study

We apply our method to DICE\textsuperscript{4}[23], a Medical TS on reasons for admission in intensive care. The DICE knowledge base, which is under development at the authors’ institution, contains about 2500 concepts. Each concept is described in both Dutch and English by a preferred term, and any synonym(s) for both languages. In addition to reasons for admission, DICE contains concepts regarding anatomy, etiology and morphology. DICE was represented using KRSS syntax\textsuperscript{4}. As an example of this syntax, the definition for Pneumonia in Figure 6.3 is represented as: 

\[(\text{define-primitive-concept Pneumonia (AND Disease \exists location Lung)})\]  

In non-primitive definitions the term define-concept is used instead of the term defineprimitive-concept.

Determining concepts of interest We have focused our evaluation on the reasons for admission taxonomy, and did not yet analyze the other taxonomies, such as anatomy and etiology. Focusing on the reasons for admission is motivated by the fact that these are the central concepts in DICE, and we want to ensure that these are defined as complete as possible. DICE contains 1456 reasons for admission.

Exclusion of poorly defined concepts The use of the KRSS syntax made it straightforward to detect all concepts that are subsumed by one concept and do not specify any difference with their subsumer. A text-based search in the KRSS file results in all definitions that contain exactly two concept names (i.e. the concept being defined and its subsumer), and no constructors (e.g. “AND” or “SOME”). These definitions are assumed to be primitive definitions. 106 concept definitions (7% of all reasons for admission) were found in this step.

\textsuperscript{4} development of DICE is supported by the National Intensive Care Evaluation (NICE) foundation.
6.4. Case Study

Tab. 6.1: Results of detection of equivalently defined concepts in the Reason for Admission module of DICE.

<table>
<thead>
<tr>
<th># equivalent definitions in set</th>
<th># sets</th>
<th># concepts in set</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>74</td>
<td>148</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>121</strong></td>
<td><strong>344</strong></td>
</tr>
</tbody>
</table>

**Assuming non-primitive definitions** The remaining 1350 concept definitions (93% of all reasons for admission) were assumed to be non-primitive.

**Inference of equivalence** RACER was used to classify the resulting terminological system. As a result of this classification it was determined that 1006 concepts (75% of the 1350 non-primitive concept definitions) had a unique definition, and 344 concepts (25%) had definitions that were logically equivalent to those of other concepts. These 344 definitions originated from 121 concept definitions that occurred twice or more, as is shown in Table 6.1. There were 74 sets of two equivalent definitions, and one set of 13 concepts with an equivalent definition.

**Interpretation of equivalence** As explained in Section 6.3, there can be various explanations for concept equivalence. Equivalent concepts need to be analyzed in order to determine whether they are actually duplicately defined, or underspecified. Such underspecification can be inevitable when concepts are natural kinds, or when concept properties can not be expressed due to limits of the used DL. Avoidable underspecification concerns tacit knowledge, which could be made explicit by enhancing concept definitions, in order to make the definitions more complete.

An in-depth description of the outcomes of analysis of equivalent definitions within DICE is beyond the scope of this paper and of limited relevance. We will here discuss only some illustrative outcomes.

Four concept definitions were found in DICE that were duplicates. These were: \{hemothorax, hemopleura\}, \{morbus Plummer, nodular toxic goiter\}, \{Guillain-Barré syndrome, inflammatory demyelinating polyradiculoneuropathy\}, and \{pneumonectomy, lung excision\}. These examples show that synonyms that are very different from each other easily remain unrecognized as such by modelers and therefore such synonyms can easily introduce duplicate
6. DL-based Auditing: Equivalence

If equivalence does not concern duplicate definitions of the same concept, it reveals concepts that differ in meaning in a way that is not represented in the knowledge base. These concepts need to be analyzed, in order to determine whether it is possible to express the distinction between them.

In DICE, a small number of natural kinds was found, which were to a large extent syndromes and/or eponyms. Examples of these are “adult respiratory distress syndrome”, and “Wolff-Parkinson-White syndrome”.

Some concepts revealed underlying semantics that could not be expressed using the representation of DICE. DICE originally had a frame-based representation, and has been migrated to DL in order to be able to perform the experiments described. A small number of concepts were found that explicitly mentioned negation, which can not normally be represented using frames. Examples of these are “bleeding” versus “non-bleeding” and “obstructive” versus “non-obstructive”. As this difference could be explicitly represented using a DL that allows for negation, it needs to be determined whether the use of a more expressive DL outweighs any increase in computational complexity.

The vast majority of concepts that were primitively defined or that were non-uniquely defined, demonstrated underspecification that seemed to be relatively easy to avoid. This means that it is possible and appropriate to extend the definitions by adding conditions. For example, equivalence of hypocalcemia and hypercalcemia can be resolved by making explicit the level involved: respectively “below normal” and “above normal”. Equivalence of hypercalcemia and hypermagnesemia is explained by the lack of specification of the involved chemical elements (respectively calcium and magnesium). These examples demonstrate required extensions to the knowledge base, as chemical elements and levels are currently not defined in the knowledge base. There were however also many concepts that can be refined using concepts and roles that are already available in the knowledge base. For example “meningococcal meningitis” was primitively defined as a “bacterial meningitis”, in spite of existence of a role “etiology” and a concept “meningococcus”. Hence, making this definition more complete is not only straightforward, it is even required, if one wants to be able to infer that meningococcal meningitis is a disease that is caused by meningococci.

6.5 Results and Lessons learned

The case study shows that there are five typical causes that result in concepts with equivalent definitions. These situations are summarized below, ordered by the increasing effort needed for overcoming the potential weaknesses discovered by equivalence. For example, truly duplicate definitions can be relatively easily fixed, whereas there is no easy fix for definitions of concepts that represent natural kinds.

1. Concepts are duplicately defined. If this is the case, all but one of the duplicate concepts can be made obsolete, and the terms attached to the
obsolete concepts can be added as synonym terms of the retained concept. For example the concept hemopleura can be made obsolete, the term “hemopleura” is then added to the concept “hemothorax” as a synonymous term.

2. Concepts can be distinguished by roles and role values that are readily present in the TS. In this case, the roles and appropriate values can be added to the concept to make the distinction explicit. For example the “etiology” “meningococcus” should be specified for the concept “meningococcal meningitis”.

3. Concepts can be distinguished by roles or role values, but these are not yet present in the TS. In this case, the appropriate roles and role values must be added to the TS, and related to the concept. For example chemical elements can be added to DICE, in order to be able to distinguish hypercalcemia from hypermagnesemia.

4. Concepts can in principle be distinguished but due to limitations of the underlying formalism this can not be expressed. In this case, a more expressive formalism can be considered, but the practical and computational consequences must be taken into account. Based on the outcome of this analysis, either a more expressive formalism can be introduced, or the limitations, which then have been clearly identified, are to be accepted. For example, the formalism of DICE can be extended with negation, in order to distinguish “bleeding” from “non-bleeding”.

5. Concepts represent natural kinds, hence their definition can not be expressed in any formal way. In this case, it is not possible to define necessary and sufficient conditions for a concept without resorting to metaphysics (e.g. introduction of the property of “dogness” to uniquely define a dog). However, an attempt can be made to define all necessary conditions (if any) in order to define the concept as precise as possible.

6.6 Discussion and Conclusions

We have devised a method for auditing medical terminological systems based on detecting concepts with equivalent definitions. We applied this method to a terminological system regarding reasons for admission in intensive care. Before drawing conclusions on the applicability of this method, we will discuss the impact of underspecification on practical use of a terminological system. Furthermore we will discuss generalizability of the results of our case study by comparing the results with an additional but preliminary case study on the Foundational Model of Anatomy [2].

6.6.1 Impact on practical use

So far we have only touched on the potential problems that duplicate definition and underspecification may cause when using a terminological system in real
practice. Now that we have discussed the possibilities of DL-based reasoning we review these problems.

**Duplicate definition** As mentioned in Section 6.1, duplicate definitions hamper querying a terminological system. If a concept is sought, any redundantly defined concepts might not be retrieved. For example, if the terminological system DICE is used to query patients with hemopleura or any of its subsumees, patients registered as having a hemothorax (or any of its subsumees) will not be returned.

**Omission of properties** Concepts can not be retrieved by means of properties that are not specified for that concept. For example, the concept “meningococcal meningitis” mentioned above will not be retrieved when querying for concepts that have “etiology” which is “meningococcus”, because this property is not expressed in the definition of “meningococcal meningitis”.

**Primitive definition** In Section 6.2.3 it was explained that inference of subsumption is blocked when concepts are represented as primitive. The example in Figures 6.1 and 6.2 demonstrated how this may lead to concepts that are not properly classified. Moreover, when a TS is used to construct concepts (i.e. post-coordination), these constructed concepts may also not be classified properly. In the example above, if a user would register a patient with an inflammation that is located in the membranes of the brain and is caused by meningococi, it should be classified as a meningococcal meningitis, but this will not occur if meningococcal meningitis is defined primitive.

The examples above show the importance of unique and complete concepts definitions, reducing the use of primitive definitions as much as possible.

### 6.6.2 Preliminary case study on FMA

An interesting question is to what extent the findings in DICE are system-specific. To this end we carried out a second case study on the FMA. FMA, developed by the University of Washington, provides about 69000 concept definitions, describing anatomical structures, shapes, and other entities, such as coordinates (left, right, etc.). The FMA Knowledge Base, which is implemented as a frame-based model in Protégé, has been migrated to DL, where specified slot-fillers in the frame-based representation were interpreted as existentially quantified roles (i.e. using the $\exists$ constructor). The resulting TS was represented using KRSS syntax. After applying the first three steps of our method, the DL-based representation of all of FMA contained about 50% primitive and 50% non-primitive concept definitions. Due to its large size we were not able to classify the full TS with RACER. We hence limited the case study to “Organs”, which comprises a convenient subset that is representative for the FMA. Of the

---


3826 concept definitions, 2659 (69%) were defined as non-primitive, and 1167 (31%) as primitive. Classification with RACER resulted in 2156 concepts (81% of the 2659 non-primitive concept definitions) with a unique definition, and 503 concept definitions (19%) that were equivalent to other definitions. These 503 definitions originated from 160 concepts that were defined twice or more. There were 106 sets of two equivalent definitions, 30 sets of three, until 1 set of 54 equivalent concepts. The distribution of numbers of sets for various sizes was comparable to the distribution found in DICE, as shown in Table 6.1.

28 sets contained concepts that referred to laterality (e.g. Left phrenic nerve, Right phrenic nerve), without explicit reference to laterality in the definition. In general, for many of the equivalent concepts, the related terms denoted positional information, e.g. distal/middle/proximal, or posterior/anterior, but this was not represented in the definition.

Hence, making concept definitions more complete using readily available concepts and roles, seemed possible in FMA. For example, “Synovial tendon sheath of flexor hallucis longus” and “Synovial tendon sheath of tibialis anterior”, can be distinguished from each other by explicitly relating them to “flexor hallucis longus”, and “tibialis anterior”, respectively.

This case study demonstrates that opportunities for further improvement can also be found in FMA, and probably in other large frame- or DL-based systems.

6.6.3 Conclusions

We have applied the inferential powers of DL reasoners to detect concepts that are equivalently defined within a knowledge base. In order to be able to find such concepts, we have considered definitions to be non-primitive. A description logic reasoner for classification of the resulting knowledge base generates sets of equivalently defined concepts.

In the literature it is hypothesized that underspecification is a common phenomenon in medical terminological systems. Our two case studies confirm this hypothesis. The vast majority of concept definitions that turn out equivalent can be improved by adding necessary conditions to the definition. Further analysis is needed to determine whether this leads to sufficient conditions as well. For some equivalent concepts there seemed to be no possibilities of improving the definition, this is because these concepts represented natural kinds, or could not be expressed due to limits of the underlying representation.

Overall, it can be concluded that the application of the method described in this paper contributes to pointing out which concepts suffer from underspecification or duplicate definition. It needs to be determined whether this method can lead to a significant decrease in the number of primitive concepts in knowledge bases, thus increasing the powers of knowledge-based inference. Although the method has been applied only to two knowledge bases in the field of medicine, it is likely that it is applicable to other domains as well.
Acknowledgments

This work has been partially funded 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.

A preliminary version of this work has been presented at the 2004 International Workshop on Description Logics - DL2004, in Whistler, Canada.

The contributions of Prof. Hasman in the preparation of the manuscript are highly appreciated.

Bibliography


