Methods for auditing medical terminological systems
Cornet, R.

Citation for published version (APA):
Summary and general discussion
The main contributions of the research presented in this thesis are the framework for characterization of medical terminological systems, and the methods for their evaluation and auditing.

Terminological systems are systems that contain terms pertaining to a certain domain. Some of these systems also represent the meaning of these terms, by providing for each term either a free-text or a formal definition.

The contribution of the framework that characterizes terminological systems is a better understanding of these systems. A good understanding of terminological systems is necessary for comparison, application, and development of these systems. Methods to evaluate content coverage, namely ‘concept matching’, ‘expert review’, and ‘formal algorithmic evaluation’ have been compared, showing that they complement each other. The auditing methods are a means to provide trust in the contents of terminological systems by assessing the consistency and completeness of concept definitions in these systems. The usefulness and usability of the methods have been demonstrated by applying them to terminological systems in the domain of medicine.

This chapter summarizes and discusses our work. It puts forward the principal findings (Section 7.1), and addresses strengths and weaknesses of our approach (Section 7.2). Furthermore, it relates our work to that of others (Section 7.3), and explains the significance of our work (Section 7.4). Finally, it provides an outlook on future research (Section 7.5) and presents concluding remarks (Section 7.6).

### 7.1 Principal Findings

In the introduction in Chapter 1, three main questions were posed that served as a guide for the research described in this thesis. We reiterate and address these main research questions below.

**How to characterize medical terminological systems?**

There are many terminological systems in medicine, which serve different purposes (e.g., recording clinical information of patients, supporting epidemiological research), and cover various domains (e.g. primary care, mental disorders). There are many ways to characterize terminological systems. One way is to describe desired characteristics thereof [1]. However, whether characteristics are desired may depend on the intended application of these systems. Therefore we suggest to describe general characteristics of systems, independent of the application. These characteristics can be regarded as consisting of a feature-value pair. For example, the characteristic “concepts are represented in Dutch and English” can be mapped onto a feature “in what languages are terms described”, and a feature value “Dutch and English”. In this way, the features of terminological systems are made explicit, and their values can be determined for each system.

The number of features extracted from the literature was large. Therefore, a systematization of these features was needed. The systematization was realized...
by introducing a two-axial categorization of the features, which distinguishes between the types of terminological systems (terminology, thesaurus, classification, vocabulary, nomenclature, and coding system) and the constituting elements of terminological systems and servers (the formalism underlying their content, the content itself, and the functionality of terminology servers). Terminology servers provide functionality for navigation, manipulation and/or modification of a terminological system, and are generally terminological-system-specific.

The framework, which was presented in Chapter 2, describes the essential features of terminological systems and servers. This framework enhances the understanding of terminological systems, which is necessary for comparison, application, and development of these systems. Since most characteristics are application-independent, their description needs to be determined only once, and can be reused for different applications of terminological systems.

Content coverage, which is a pivotal feature of terminological systems, was found to be the only feature in the literature that could not be represented in an application-independent way. Various content-coverage measures are made explicit in Chapter 3 and methods to determine these measures are presented and applied. “Term coverage” and “concept coverage” can be measured with respect to samples representative for the intended use (e.g., recording patient information, performing scientific medical research). In addition to these term and concept coverage methods, methods for manual and automated evaluation of the contents are presented and assessed, and a comparison is made of methods to evaluate the content of terminological systems.

The main finding of Chapter 3 was that the various methods presented are complementary. An assessment of how well the domain of use is covered can best be performed using concept matching methods. Completeness and correctness of definitions can best be assessed by experts or by the logic-based automated methods. Currently, expert review seems to outperform the automated methods, but at the cost of a much higher effort.

How to exploit the benefits of formal representation for auditing methods?

An increasing number of terminological systems represent definitions of concepts in a formal way, using either frames [2] or description logic [3]. As these formal definitions can be used for various purposes, e.g., decision support or aggregation of patient groups, it is essential that these definitions are correct and as specific as possible. Reasoning based on description logic addresses inconsistent definitions and equivalent definitions, and is therefore considered as a means to facilitate the auditing process. We devised a method to migrate from frame-based to description-logic-based representations. Our research shows that the contribution of description-logic-based representation to auditing the contents of terminological systems depends on the way the contents are modeled. Most notably, the use of closure axioms (i.e., disjointness statements and universal restrictions) and the assumption of equivalence contribute to auditing of the contents with respect to logical consistency and completeness of concept definitions. Furthermore, we have developed pinpointing techniques that help
7.2. Strengths and Weaknesses of our Approach

In our approach to the development of methods for auditing terminological systems we strived for using methods that are as much as possible generally applicable and implementable. First, this concerns being applicable to a broad range of terminological systems, rather than being aimed at a single system. Second, it involves the tools needed to perform auditing, such as the description-logic reasoners. The use of description logic and of standard inference keeps the auditing methods independent from any specific tools.

As the framework aims to be general, its completeness with regard to characteristics that are specific to certain kinds of terminological systems may be limited. For example, depending on the representation language, additional
characteristics may be of interest. Such characteristics are currently not yet covered, but can be added, as the framework can easily be extended. The framework has so far only been applied to describe SNOMED CT, but in order to assess its full potential and determine the completeness of the characteristics covered, more terminological systems would have to be characterized by means of the framework.

There are two major advantages of applying description-logic reasoning for auditing terminological systems. First, the use of well-studied mechanisms and off-the-shelf reasoners increases understandability and reproducibility of the methods and their results. Second, description-logic-based representation is a driving force for making the semantics of concept definitions explicit.

A drawback of the description-logic-based reasoners is that they were not able to deal with very large terminological systems, but only smaller-sized parts thereof. This may lead to missing inconsistencies which exist in the system as a whole but are not present in the audited parts of the system, for example when two conflicting axioms end up in different parts. In addition, although the methods do provide a straightforward strategy that results in potentially many erroneous axioms, there is no way to focus on certain types of errors, for example errors in specific relationships (e.g., \textit{hasEtiology}). As a result, processing the outcomes of the methods is labor-intensive.

In our approach the application of the methods was performed as a separate activity. In practice however, auditing should be a part of the modeling process, hence the methods should be integrated in this process, which requires dedicated software tools.

The methods were applied to a real-world terminological system, DICE. Application of the methods to other systems may show a significantly different number of potentially or actually erroneous concepts found. A preliminary study [5] on the Foundational Model of Anatomy (FMA) [6] showed that the proportion of erroneously defined concepts is smaller in FMA, but that the proportion of underspecified concepts is comparable to that of DICE.

The method to detect incorrect definitions is limited as incorrect definitions that do not conflict with other definitions remain undetected. So far, the proportion of incorrect definitions that are revealed by using this method has not been determined.

Our methods aim at assessing and increasing the quality of terminological systems, but we did not determine how this increase in quality is related to specific errors, or types of usage of the systems, such as epidemiological research. Consequently, it is yet unclear what the effect of auditing is on practical use. For example, errors of concepts that refer to frequently occurring diseases might be more important than those which rarely occur.

7.3 Related Research

The research presented in this thesis – the framework, the content-evaluation methods and the auditing methods – builds on research in the domains of med-
ical informatics and computer science.

The framework is to a large extent indebted to key publications on terminological systems [1, 7–9]. Features presented in the framework were distilled from these publications, which thus form the foundations of the framework.

The methods to evaluate contents of terminological systems as described and compared in Chapter 3 were extracted from over 20 studies on content coverage of terminological systems performed in the last decade, e.g., [10–13]. For these studies, the focal points of evaluation were made explicit and compared.

Studies on auditing increasingly take into account the fact that many terminological systems explicitly represent terms and concepts. Terms and concepts correspond to “symbol” and “thought or reference” respectively of the semiotic triangle depicted in Figure 1.2. Auditing methods presented in the literature either focus on terms, or focus on concepts or use a hybrid approach, combining the lexical information from terms with the hierarchical ordering of concepts. On the “symbol” side of the spectrum is [14], focusing on missed synonymy. On the “thought or reference” side of auditing are [15–17], that focus on correctness of the concept hierarchy. Hybrid approaches are described in [13] and [18], where lexical matching and synonym recognition are combined with information from the concept hierarchy.

The auditing methods that were introduced in Chapter 4 and 6 do not pay attention to the terms that designate concepts. The focus is on concepts, more precisely the definitions of concepts. This has sofar been an underexposed research area. Previous research addressing concepts in terminological systems mainly involves hierarchical placement, but not the correctness and completeness of the definitions of the concepts.

Applying description logics in the domain of medicine increasingly got attention [19–23] since the GALEN project, that started in the early 1990s [24]. Since then, research focuses on possibilities of how to represent concept definitions using description logic, mainly in order to benefit from the reasoning capabilities offered, as described in [3]. Application of reasoning is extended in our research by the use of closure axioms and equivalence axioms in a description-logic-based representation to support detection of inconsistently defined or underspecified concepts.

Our methods did prove useful for auditing, but they cannot replace existing methods, as these focus on other aspects, and may use other information from the terminological systems, for example lexical information. Hence, our methods are essentially an addition to existing methods. Whereas our auditing methods have been related to other research, the methods were not compared with regard to their efficiency and effectiveness. Consequently, it can not be determined which methods are the most suitable for a specific terminological system, or for a specific type of analysis that is required.
7.4 Significance of the Research

The number of terminological systems that provide formal concept definitions is increasing. Such systems generally represent concept definitions using frames or description logics. For these systems, auditing of the definitions can be performed using the methods described in this thesis. The foremost motivation for auditing is to provide trust in medical terminological systems. This is important for knowledge modelers who are developing and maintaining terminological systems, for clinicians who use them in everyday clinical practice, and for clinical/epidemiological researchers who use them to aggregate groups of patients. The relevance of auditing to these groups is discussed below.

As the number and size of terminological systems increases, it is important for knowledge modelers to be able to properly characterize the systems in order to perform analyses on strengths and weaknesses thereof, which is facilitated by the framework described in Chapter 2. For example, if the characterization reveals that hierarchical placement of concepts is restricted in breadth or depth, knowledge modelers either have to cope with this limitation when adding concepts to the system, or the system has to be altered in order to overcome this limitation.

Another benefit comes from the methods that provide means for auditing the contents of terminological systems. Knowledge modeling is a very labor-intensive effort that is generally characterized by involvement of a team of knowledge modelers and domain experts, spanning a larger period of time. This is well illustrated by SNOMED CT, which has a 40-year history since its first predecessor, SNOP, that stems from 1965. Its editorial board consists of about 30 people. Currently, SNOMED CT contains over 300,000 concepts. These figures make clear that automation of the knowledge modeling process is instrumental. The auditing methods of applying description-logic-based representation with closure axioms and equivalence axioms can contribute to this automation, providing a structured means for evaluation of the represented knowledge in terminological systems. As the methods shift the burden of the auditing process from human to machine, the effort of knowledge modelers is reduced.

When terminological systems are to be used for registration of patient information, good coverage of the domain of use is crucial. The methods presented in Chapter 3 contribute to determining the coverage of terminological systems, in order to assess the suitability of a system or, when more systems are available, compare these systems. For other use, such as automated decision support, correctness of the definitions is very important, in order not to miss or make incorrect inferences. The methods presented in this thesis can be applied to evaluate terminological systems with regard to this purpose.

Whether terminological systems can be trusted to be used in medical research heavily relies on the quality of the concept definitions. For medical research purposes, aggregation of patient groups plays an important role. Such aggregation can be based on the (pre-coordinated) hierarchical ordering of concepts, but can also be constructed ad-hoc (post-coordinated), based on the definitions of concepts. Ad-hoc construction of aggregates relies on the correct-
ness and completeness of concept definitions, which can be determined by the auditing methods proposed in this thesis.

Our work furthers the field of research on terminological systems by standardization of ways to characterize, evaluate and audit these systems. Furthermore it broadens the scope of application of description-logic-based reasoning to real world problems. It also demonstrates the added value of closure axioms and equivalence axioms, and introduces new ways to pinpoint potential modeling errors.

7.5 Further Work

The emphasis of the research has been on the development of a framework and auditing methods. Further research involves the methods, their application, and further application of description logics.

Further research on the existing methods involves integration of the methods in the modeling process. Sofar, auditing has been performed independently of the modeling process, but an integrated approach will contribute to quicker error detection and correction. Additionally, a comparison between existing auditing methods and those presented in this thesis is required in order to gain insight into their effectiveness and efficiency. Ultimately, the auditing methods and the pinpointing methods should be combined and applied to larger, well-known medical terminological systems.

Further application of the methods to other terminological systems will prove useful for assessing the merits of these methods in real practice, and shed light on possible improvements to and extensions of the methods. Assessment is needed of the practical value of the methods for the use of terminological systems in clinical practice and for clinical or epidemiological research purposes, in order to determine the importance of using the auditing methods, other than providing trust in these systems.

Further application of description logics involves using description logics not only in auditing, but also in practical use of medical terminological systems. Issues to be studied further are how to store (patient) information that was recorded using description-logic-based terminological systems, and possibilities for reasoning with this information, most notably for the purpose of aggregating patients into groups.

7.6 Concluding Remarks

The work described in this thesis has been performed in the context of a larger project, entitled “Making semantics explicit: supporting terminology for Computer Patient Record users”. At the start of the project, many issues regarding terminological systems were unclear, such as their structure, their contents, and how these could be described and evaluated. The methods we have presented were the result of our quest for further understanding of terminological systems, that started with the research of De Keizer et al. [25, 26].
Understanding of terminological systems is essential to fully exploit their strengths. We believe that our methods contribute to furthering this understanding, and to providing trust in terminological systems. We hope that this will pave the way for successful use of terminological systems in health care, which will contribute to computer patient records containing understandable and unambiguous patient information. This information will be vitally important for clinical and epidemiological research, as well as for further improvement of patient care.

Bibliography


