Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

NJIT
New Jersey's Science & Technology University

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

EXTENSIONS OF SNOMED TAXONOMY ABSTRACTION NETWORKS SUPPORTING AUDITING AND COMPLEXITY ANALYSIS

by
Duo Wei

The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) has been widely used as a standard terminology in various biomedical domains. The enhancement of the quality of SNOMED contributes to the improvement of the medical systems that it supports.

In previous work, the Structural Analysis of Biomedical Ontologies Center (SABOC) team has defined the partial-area taxonomy, a hierarchical abstraction network consisting of units called partial-areas. Each partial-area comprises a set of SNOMED concepts exhibiting a particular relationship structure and being distinguished by a unique root concept. In this dissertation, some extensions and applications of the taxonomy framework are considered. Some concepts appearing in multiple partial-areas have been designated as “complex,” due to the fact that they constitute a tangled portion of a hierarchy and can be obstacles to users trying to gain an understanding of the hierarchy’s content. A methodology for partitioning the entire collection of these so-called overlapping complex concepts into singly-rooted groups was presented. A novel auditing methodology based on an enhanced abstraction network is described.

In addition, the existing abstraction network relies heavily on the structure of the outgoing relationships of the concepts. But some of SNOMED hierarchies (or subhierarchies) serve only as targets of relationships, with few or no outgoing relationships of their own. This situation impedes the applicability of the abstraction network. To deal with this problem, a variation of the above abstraction network, called the converse abstraction network (CAN) is defined and derived automatically from a given SNOMED hierarchy. An auditing methodology based on the CAN is formulated.

Furthermore, a preliminary study of the complementary use of the abstraction network in description logic (DL) for quality assurance purposes pertaining to SNOMED is presented.

Two complexity measures, a structural complexity measure and a hierarchical complexity measure, based on the abstraction network are introduced to quantify the complexity of a SNOMED hierarchy. An extension of the two measures is also utilized specifically to track the complexity of the versions of the SNOMED hierarchies before and after a sequence of auditing processes.

EXTENSIONS OF SNOMED TAXONOMY ABSTRACTION NETWORKS SUPPORTING AUDITING AND COMPLEXITY ANALYSIS

by
Duo Wei

A Dissertation
Submitted to the Faculty of the
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science

August 2011

Copyright © 2011 by Duo Wei

ALL RIGHTS RESERVED

APPROVAL PAGE

EXTENSIONS OF SNOMED TAXONOMY ABSTRACTION NETWORKS SUPPORTING AUDITING AND COMPLEXITY ANALYSIS

Duo Wei

Dr. Yehoshua Perl, Dissertation Co-Advisor Date
Professor, Computer Science Department, NJIT

Dr. Michael Halper, Dissertation Co-Advisor Date
Director, Information Technology Program, NJIT

Dr. James Geller, Committee Member Date
Professor, Computer Science Department, NJIT

Dr. Gai Elhanan, Committee Member Date
Research Professor, Computer Science Department, NJIT

Dr. Chunhua Weng, Committee Member Date
Assistant Professor, Biomedical Informatics Department, Columbia University

BIOGRAPHICAL SKETCH

Author: Duo Wei

Degree: Doctor of Philosophy

Date: August 2011

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, 2011

- Bachelor of Science in Computer Science,
  Changchun Normal University, Changchun, P. R. China, 2006

Major: Computer Science

Presentations and Publications:

D. Wei, Y. Perl, M. Halper. Structural complexity measures to track the impact of auditing on SNOMED hierarchies, to be submitted for journal publication, 2011.

D. Wei, M. Halper, G. Elhanan. Using SNOMED semantic concept groupings to enhance semantic-type assignments in the UMLS, submitted to *2nd ACM SIGHIT International Health Informatics Symposium*, 2011.

Y. Wang, M. Halper, D. Wei, Y. Perl, and J. Geller. Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED, *Journal of Biomedical Informatics*, accepted.

Y. Wang, M. Halper, D. Wei, H. Gu, Y. Perl, J. Xu, G. Elhanan, Y. Chen, K. A. Spackman, J. T. Case, G. Hripcsak. Auditing complex concepts of SNOMED using a refined hierarchical abstraction network, submitted to *Journal of Biomedical Informatics*, 2011.

D. Wei, O. Bodenreider. Using the abstraction network in complement to description logics for quality assurance in biomedical terminologies - a case study in SNOMED CT. *Medinfo*, pp.1070-1074, Cape Town, South Africa, Sep. 2010.

D. Wei, M. Halper, G. Elhanan, Y. Chen, Y. Perl, J. Geller, and K. A. Spackman.

Auditing SNOMED relationships using a converse abstraction network, *Proceedings of the 2009 American Medical Informatics Association Annual Symposium*, pp. 685-689, San Francisco, CA, Nov. 2009.

H. Gu., D. Wei, J. Mejino, and G. Elhanan. Relationship auditing of the FMA ontology, *Journal of Biomedical Informatics*, vol. 42, no. 3, pp. 550-557, Jun. 2009.

Y. Wang, D. Wei, J. Xu, G. Elhanan, Y. Perl, M. Halper, Y. Chen, K. A. Spackman, and G. Hripcsak. Auditing complex concepts in overlapping subsets of SNOMED, *Proceedings of the 2008 American Medical Informatics Association Annual Symposium*, pp. 273-277, Washington, DC, Nov. 2008.

D. Wei, Y. Wang, Y. Perl, J. Xu, M. Halper, and K. A. Spackman, Complexity measures to track the evolution of a SNOMED hierarchy, *Proceedings of the 2008 American Medical Informatics Association Annual Symposium*, pp. 778-782, Washington, DC, Nov. 2008.

To my family,
for their endless love and support.

To Thomas,
my friend, my love, and my husband.

ACKNOWLEDGMENT

The tremendous effort and good will of many people have enabled completion of this dissertation. I am very thankful for all the help I have received throughout my PhD studies. I really appreciate that I have such a chance to express my appreciation and respect to these people.

First of all, I am full of gratitude to my dissertation advisors Dr. Yehoshua Perl and Dr. Michael Halper for their great help, encouragement and guidance in all aspects of my research. I really appreciate that they offered me such a wonderful opportunity to study medical informatics, and no words could fully encompass the amount of gratitude I have for their supervision and support. Their broad knowledge, great expertise in research and kind personality has affected me deeply, which have benefited me greatly in the past and will be of important value in my future career. I really feel honored to be their student.

I am heartily grateful to Dr. James Geller for his generous help and valuable advice in my research. I also thank him for his effort for serving as a committee member of my dissertation defense. His rigorous approach, attitude and conscientiousness in research have impressed me deeply. His sense of humor and vivid teaching style benefit me continually in the future.

I would like to thank Dr. Gai Elhanan and Dr. Chunhua Weng for their advice and help effort in my graduate study and dissertation defense.

I would like to thank my friends in Structural Analysis of Biomedical Ontologies Center (SABOC). They are Huanying Gu, Charles Paul Morrey, Kuo-chuan Huang, Yue

Wang, Yan Chen, Zhe He, Ankur Agrawal, Christopher Ochs, David Daudelin, and all other people in the group. I appreciate their help and enjoy the time with them.

Special thanks should be given to Dr. Olivier Bodenreider, the staff scientist of the National Institutes of Health (NIH), for his generous support and kindly guidance during my summer internship at NIH in 2009.

Finally, and most importantly, I would like to thank my parents, Kening Wei and Shurong Liu, for their endless love, encouragement and support during my life, even in those hard days. Their kindness and diligence will impact me throughout my life. My gratitude to them can never be overstated. Thank you to my wonderful husband, Gang, for being patient, supportive, and helping at every stage of my life. You never let me feel alone, even when we are distanced. You are my soul mate!

The research work presented in this dissertation is supported by the National Library of Medicine under grant R-01-LM008912-01A1.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

With the advent of electronic health record (EHR) systems, IT solutions are needed to simplify the recording of standard codes for clinical providers as well as practitioners. The basis for these products is a standard terminology, without which the full benefits of an EHR are unlikely to be realized.

The Systematized Nomenclature of Medicine – Clinical Terms ("SNOMED" for short, hereafter), one of the most popular standard terminologies, is well structured, highly computerized, and has many merits that make it superior to its peers. Among its advantages are: (1) the contents are systematically organized, which allow a consistent way of indexing, storage, retrieval, and integration; (2) concepts in the SNOMED have clear definitions that are unambiguous, where each concept has a Fully Specified Name (FSN), a preferred term, and potentially several synonyms. This helps to avoid ambiguity; (3) concepts are context-free and post-coordination is encouraged, which means a piece of clinical text can use more than one code or attribute to explain it; (4) Description Logic (DL) based modeling, which allows for DL-based classifiers that can position concepts within hierarchies and identify inconsistencies in the content. With all these features, concepts in the SNOMED cover a variety of domains, such as clinical finding, procedure, substance, and events.

However, due to SNOMED's large number of concepts as well as the complicated network of relationships among them, it is inevitable that errors will find their way to this large knowledge base. Either a modeling error or a content error can result in very serious adverse effects. For example, according to statistics, more than 100,000 people die annually because of adverse drug reactions, some of which may be prevented by appropriate modeling of the EHR systems and their use of a terminology.

Because of the importance of a terminology, quality assurance of SNOMED is critical. As people's knowledge of health and healthcare is constantly evolving, the number of medical terminologies has grown tremendously; thus, the quality of SNOMED has become increasingly of significance. The International Health Terminology Standard and Organization (IHTSDO) [1] in fact formed a separate quality assurance group. It is in that group that SNOMED's content undergoes a clinical quality assurance process prior to each release.

The dissertation work is to assist in the quality assurance of SNOMED by applying computer science techniques combined with medical domain experts' review. This process is called semi-automatic auditing. In previous research, the SABOC team has devised high-level abstraction networks based on analyses of a SNOMED hierarchy's attribute relationships and their patterns of inheritance [2]. A hierarchy's concepts were partitioned into groups, called *areas*, according to their specific attribute relationships. From this partition, an abstraction network, referred to as the *area taxonomy*, affording a summary view of the distribution of the attribute relationships was constructed. Further refinement of areas led to another abstraction network, the *partial-area taxonomy*, which conveyed information about sub-area hierarchical arrangements. In addition to their support for orientation to and comprehension of a SNOMED hierarchy, the two networks have served as the bases of the formulation of structural methodologies for auditing

SNOMED hierarchies[2]. Importantly, the SABOC team has found that many concept errors manifested themselves as structural anomalies at the taxonomy level, and thus the taxonomies proved to be effective building blocks for automated auditing regimens.

The objective of the current research is to present innovative auditing methodologies that take advantage of an enhanced abstraction network for quality-assurance purposes. Moreover, the connection between the quality-assurance results and various SNOMED complexity measures defined using abstraction networks is explored.

The ultimate goal, as stated by Alan Rector from the University of Manchester, UK, is: “We will know we have succeeded when clinical terminologies in software are used and re-used, and when multiple *independently developed* medical records, decision support, and clinical information retrieval systems sharing the same information using the same terminology are in routine use” [3].

1.2 Background and Literature Review

1.2.1 Biomedical Terminologies and Their Application

Biomedical informatics is an active research field, with the terminology sub-field gaining a lot of attention. Biomedical terminologies are critical for integration of data from diverse sources and for use by knowledge-based biomedical applications, especially natural language processing and associated mining and reasoning systems.

A terminology is a repository of concepts pertaining to topics such as diseases, primary care, procedures, diagnostics, disorders, genes, laboratory observations, etc. In 1986, Donald Lindberg and Betsy Humphreys from the National Library of Medicine (NLM) launched a project to construct a resource that would bring various disseminated controlled medical terminologies [4-6] together to form the so-called *Unified Medical Language System* [7-9].

Standard biomedical terminologies and ontologies have evolved significantly over the past one hundred years or so. In 1893, one of the best known terminologies, the International Classification of Diseases (ICD) [10], was first published to classify diseases and a wide variety of signs, symptoms, abnormal findings, and social circumstances. In 1977, its Ninth Edition (ICD-9) was published by the World Health Organization (WHO), and 15 years later (in 1992), WHO published the Tenth Edition (ICD-10). In 1978, the US national Center for Health Statistics published the ICD-9 with Clinical Modifications (ICD-9-CM) [10]. ICD-9-CM has been widely adopted in the US and around the world since then. Subsequently, many other medical terminology systems emerged to meet different demands. These include the International Classification of Primary Care (ICPC) [11], Current Procedural Terminologies (CPT) [12], Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [13], the Gene Ontology (GO) [14], Logical Observations, Identifiers, Names, and Codes (LOINC) [15-17], the Medical Subject Headings (MeSH) [18], the Foundational Model of Anatomy (FMA) [19], and the UMLS [7-9].

Biomedical terminologies exist to serve various biomedical purposes in both laboratory-based research and in actual clinical settings. First and foremost is the demands placed upon them by the pursuit of the Electronic Health Record (EHR). According to the Health Information Management Systems Society’s definition:

> *The Electronic Health Record (EHR) is a longitudinal electronic record of patient health information generated by one or more encounters in any care*

> *delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter, as well as supporting other care-related activities directly or indirectly via interface – including evidence-based decision support, quality management, and outcomes reporting*.

Biomedical terminologies play a strategic role in providing access to computerized health information because clinicians use a variety of terms for the same concept. For example, "Cardiac disorder," "Cardiopathy," or "Heart disease" might be written in the patient record – usually these are synonyms. Without a structured vocabulary, an automated system will not recognize these terms as being equivalent.

Not only are terminologies applied in the flourishing EHR systems, they are also extensively utilized in many research areas, such as knowledge management (including indexing and retrieval of data and information, mapping among ontologies); data integration, exchange, and semantic interoperability; and decision-support and reasoning (including data selection and aggregation, natural language processing applications, knowledge discovery) [20].

One example of knowledge management of biomedical terminologies is the utilization of the MeSH and UMLS as the backend search engine [21-24] to explore the biomedical literature. Besides MeSH and the UMLS, SNOMED CT is also used in a system that helps patients find physicians with particular expertise [25]. With the indexing techniques, large document collections such as MEDLINE provide high recall and high precision.

Applications of biomedical terminologies are also demonstrated in data integration and semantic interoperability process. For example, RxNorm [20], UMLS [26], and SNOMED [27] are used to exchange medication data between the Department of Veterans Affairs (VA) and the Department of Defense's (DoD's) clinical information systems. LOINC [16], in conjunction with HL7, is widely used in the exchange of laboratory data.

Another example application of terminologies is in the field of medical language processing and natural language processing [28-30]. Linguistic approaches have been developed to represent patient data. Clinical statements are classified by the biomedical terminology's categories using medical language processing systems to convert narrative information into relational database tables of patient information. During the process, clinical narratives are mapped to standard terminologies via medical language processing systems. On the other hand, terminologies (ontologies) provide the basis for free-context data mining and text mining.

In addition, some tools for domain modeling and knowledge-based representations were developed to facilitate the application of terminologies. For instance, Protégé [31-36], a free, open-source software tool, provides a growing number of users a platform to support the creation, visualization, and manipulation of ontologies in various formats. Furthermore, Protégé can be extended by various plug-ins for building knowledge-based tools and applications, such as reasoners. In this sense, Protégé is considered not just a navigation and modeling tool, but also a knowledge discovery and decision-support system.

1.2.2 SNOMED

The SNOMED [37-40] is a comprehensive clinical terminology that provides clinical content and expressivity for clinical documentation and reporting. It can be used to code, retrieve, and analyze clinical data. It was developed as a joint venture between the College of American Pathologist (CAP) and the UK's National Health Service (NHS). SNOMED was formed by merging, expanding, and restructuring an earlier version (i.e., SNOMED RT) and the UK's Clinical Term Version 3 (CTV3). In 2007, the SNOMED intellectual property rights were transferred from the CAP to the IHTSDO.

The basic building blocks of the SNOMED are concepts, descriptions, and relationships. SNOMED's concepts are organized in 19 top-level hierarchies (as of the July '09 release), each with a unique root called a top-level concept, such as Procedure, Clinical Finding, and Body Structure. This allows very detailed ("granular") clinical data to be recorded and later accessed or aggregated at a more general level. Above all these top-level concepts sits a single concept called "SNOMED CT Concept," which serves as the root of the entire terminology. Each concept is a descendant of SNOMED CT Concept via a sequence of IS-A (subsumption) relationships passing through exactly one top-level concept.

Each SNOMED concept (the anchor of meaning) is represented by a unique human readable Fully Specified Name (FSN), which is a phrase that describes the concept in a way that is intended to be unambiguous. The concepts are formally defined in terms of their relationships with other concepts. These "logical definitions" give explicit meaning that a computer can process and query on. Every concept also has a set of terms that name the concept in a human-readable way. The SNOMED includes more than 311,000 unique concepts (including inactive concepts). Examples include: Fever, X-ray of left ankle, Exposure to toxin, Family history of ear disorder, and so on.

Descriptions are the terms or names, assigned to each of SNOMED's concepts. A given concept has one or more associated descriptions. In addition to the Fully Specified Name, there is a "Preferred term" that is meant to capture the common word or phrase used by clinicians to name the concept. Many concepts have alternative descriptions called "synonyms." There are almost 800,000 descriptions in SNOMED. For example, heart disease (disorder) is the Fully Specified Name and "disorder" in parentheses is called a "semantic tag." The preferred term is Heart disease. There are several synonyms, such as Cardiac disorder, Cardiopathy, Disorder of heart, and Morbus cordis.

Relationships are the connections between concepts in SNOMED, with every concept having at least one relationship to another concept. Relationships in SNOMED are unidirectional, extending from a source concept to a target concept. Inverse relationships (from target to source) are not maintained. IS-A relationships form the basis of the hierarchies. Each connects a more specific concept (a child) to a more general concept (a parent). Additional attribute relationships characterize and further define concepts. Each can take on values (targets) only from a prescribed top-level hierarchy. IS-A relationships and attribute relationships are known as the "defining characteristics" of SNOMED concepts. For brevity, the "attribute relationship" will be referred to as "relationship," while "IS-A relationship" will be referred to as "IS-A." For example, *Bone fracture* IS-A *Bone injury*. One of its relationships *associated morphology* has a value Fracture, and the other relationship *finding site* has a value Bone structure. In total, there are approximately 1,360,000 links between the concepts.

Relationships in SNOMED are SABOC team's major interests when applying the partitioning techniques and constructing abstraction networks for auditing. Some hierarchies introduce many relationships. For example, the Procedure hierarchy introduces 23 different relationships, while some hierarchies have few relationships, such as the Physical Object, Substance, and Organism.

A stated definition view (stated view) is the set of relationships (and groups of relationships) that an author has stated to be defining characteristics of a concept. An inferred definition view (inferred view) is derived from the stated concept definition by applying a consistent set of logical rules to the definition taking account of the definitions of related concepts. The standard SNOMED CT distribution includes the relationships table that represents an inferred view of the definitions of all active concepts, restricted to the proximal super types for each concept. From January 2010, the SNOMED distribution includes the stated view, too.

1.2.3 Area Taxonomy and Partial-Area Taxonomy

In previous work, structural analyses of SNOMED hierarchies have been carried out yielding two types of high-level abstraction networks: the *area taxonomy* and the *partial-area taxonomy* [2]. Each serves to capture the relationship distribution within a hierarchy from a high-level perspective. Both networks are derived based on the respective relationships exhibited by the concepts in the hierarchy. The latter network refines the former by including additional hierarchical grouping knowledge. In the following, the important details pertaining to these two networks are presented.

The basis of the area taxonomy is a partition of the concepts into what is called *areas* according to their sets of (non-hierarchical) relationships. Five example areas are shown in Figure 1.1, where the boxes represent areas. Each area is named by placing in braces the set of relationships common to all its concepts. For example, the area {*morphology*} in the upper left has the relationship *morphology*. Concepts in the same area are listed in an indented format according to their IS-A relationships. In {*morphology*}, *Specimen from wound* IS-A *Lesion sample*, while *Specimen from wound abscess* IS-A *Specimen from wound*. On the other hand, a selected few concepts from different areas are explicitly linked by arrows to show additional IS-As. For example, *Abscess swab* in the area {*procedure*, *morphology*} IS-A *Specimen from abscess* in the area {*morphology*} and also IS-A *Swab* in the area {*procedure*}. A given concept belongs to only one area as determined by its set of relationships. Therefore, the areas taken together as a collection form a partition of a hierarchy's concepts. Five different relationships are introduced to the concepts of the Specimen hierarchy; they are *substance*, *morphology*, *procedure*, *topography*, and *identity*. Different combinations of these relationships can form different areas. The two-relationship area {*procedure*, *morphology*} and {*identity*, *procedure*} can be seen in Figure 1.1.

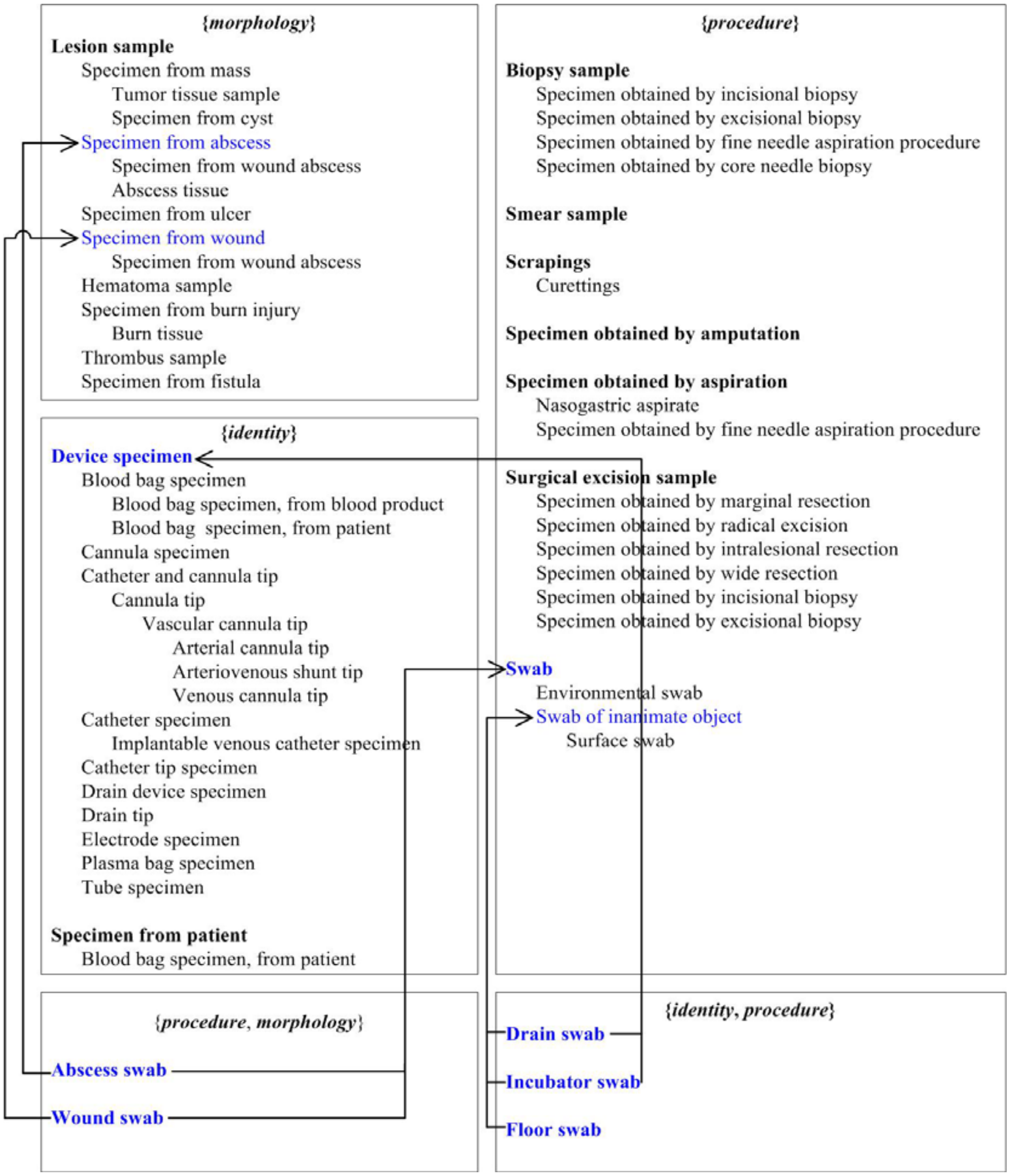


Figure 1.1 Portion of areas of the Specimen hierarchy at the concept level.

The areas are abstracted to form a network called the *area taxonomy*. The area taxonomy is a directed acyclic graph (DAG) constructed by making each area a node and then arranging them hierarchically – analogously to the underlying concepts – using what is referred to as *child-of* relationships as edges. The *child-of*'s are derived from the concepts' IS-A links. Figure 1.2 shows an excerpt of the area taxonomy of SNOMED's Specimen hierarchy (July 2007 version).

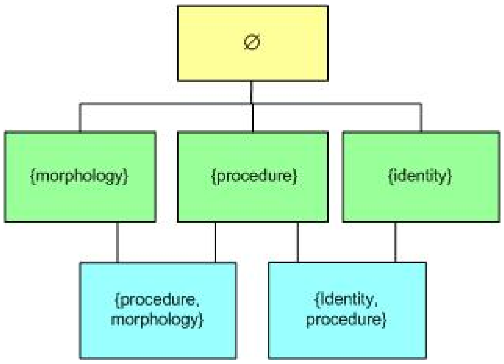


Figure 1.2 Portion of the area taxonomy of Specimen hierarchy.

A *child-of* link in the area taxonomy is derived as follows. Let A and B be two areas such that a root of A has a parent in B. Then there exists a *child-of* from A to B. Overall, the collection of area nodes and child-of edges forms a DAG. The color coding of areas is according to levels, where each level has a specific number of relationships. On Level 1, each area is labeled with one relationship; on Level 2, the areas are labeled with two relationships, etc. The yellow box at the top marked Ø means that concepts in that area have no relationships at all. So, overall, this excerpt contains six areas distributed in three different levels. On Level 1, the three areas are {*morphology*}, {*procedure*}, and {*identity*} highlighted in green. On Level 2, displayed in blue, the two areas {*procedure*, *morphology*} and {*identity*, *procedure*} are found. Unlike in Figure 1.1, no concepts are shown in the area taxonomy of Figure 1.2. This highly abstracted network provides a compact view of the relationship structure of a hierarchy.

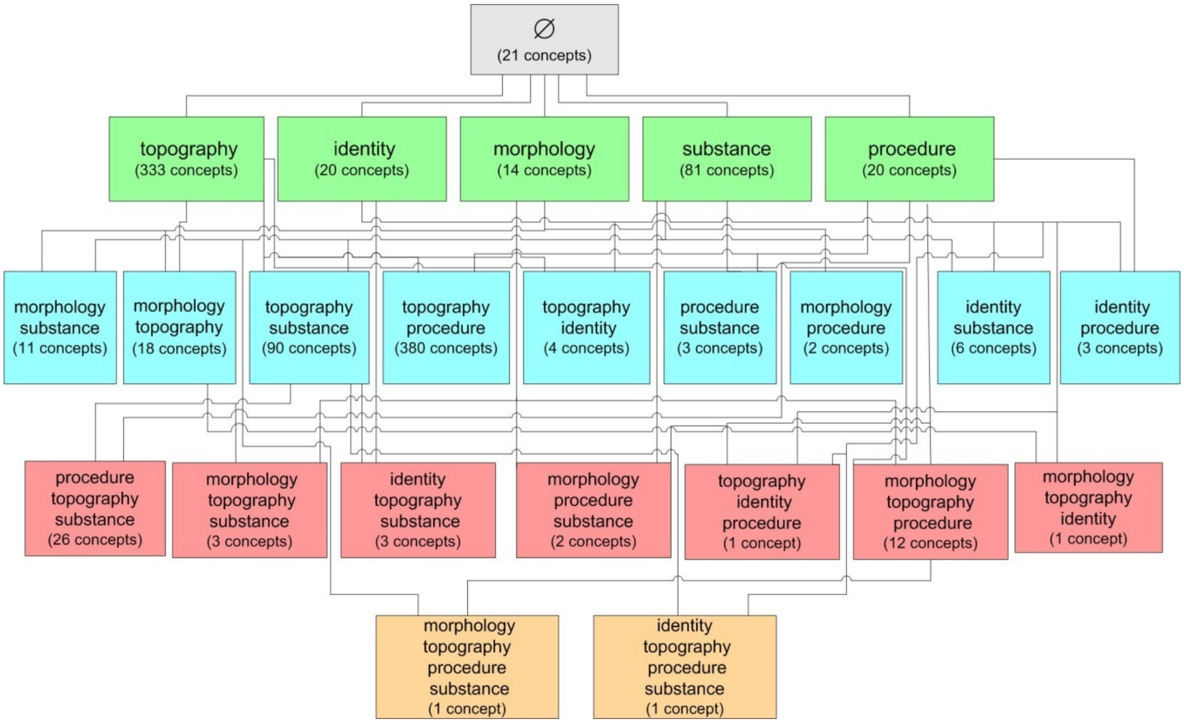
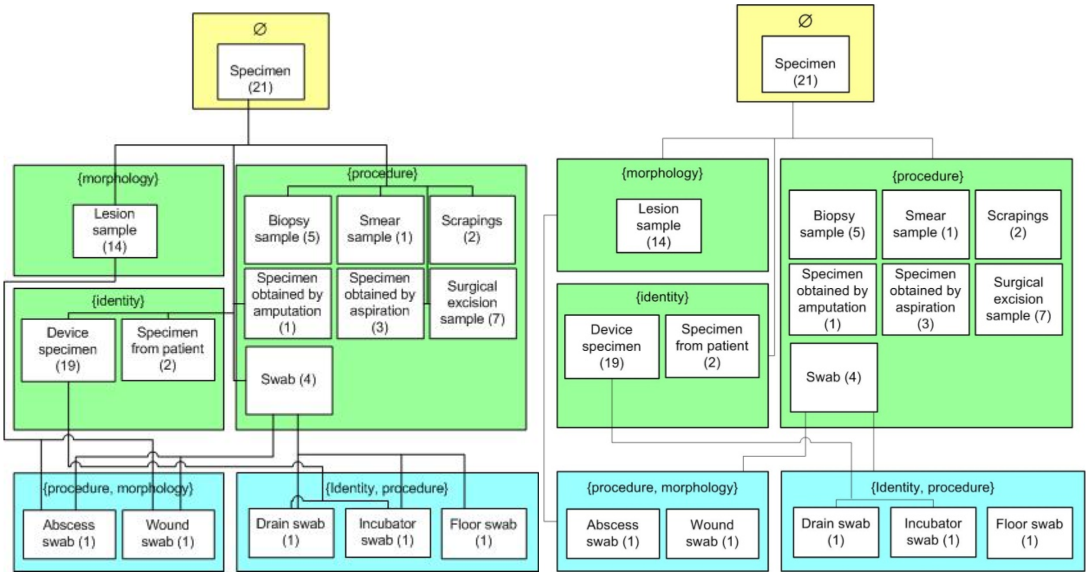


Figure 1.3 Area taxonomy for SNOMED's Specimen hierarchy.

The area taxonomy of the 1,056-concept Specimen hierarchy (July '07 release) has a total of 24 areas distributed over five levels (Figure 1.3). As stated above, the boxes are the areas and the edges (directed upward) are the *child-of*'s. The area Ø is on Level 0 and has zero relationships. The five green rectangles on Level 1 are the areas having exactly one relationship each.

Another level of abstraction is provided by the *partial-area taxonomy*. Before introducing it, let us start with the notions of root and partial area. A root of an area is a concept having no parents residing in its area. For instance, *Abscess swab* in the area {*procedure*, *morphology*} mentioned earlier is a root because one of its parents is in the area {*procedure*} and the other is in {*morphology*}. All the concepts with bold in Figure 1.3 are roots. The roots are important for an area because all other concepts are subsumed by them, and thus the roots serve as summarizations of the area's essence.



(a) (b)

Figure 1.4 Portion of the partial-area taxonomy of Specimen hierarchy.

The roots and all their descendants are further grouped into what are called partial-areas. An area may have more than one root, so it may have more than one partial-area. As is shown in Figure 1.1, the area {*morphology*} has only one root and has only one partial-area. On the other hand, {*procedure*} has seven roots and, consequently, seven partial-areas. Partial-areas form a semantic division of an area, and abstracting them as nodes leads to a network called the *partial-area taxonomy*. Figure 1.4 shows a portion of the partial-area taxonomy of the Specimen hierarchy. The names of the partial-areas are the root names and the numbers in the parentheses are the numbers of the concepts residing in the respective partial-areas (including the roots). For instance, one of the roots in {*procedure*} is called *Biopsy sample*, so there is such a partial-area. *Biopsy sample* has four descendants in its area, so the total number of concepts in its partial-area is five.

In Figure 1.4(a), the *child-of* links among partial-areas of the partial-area taxonomy are defined in a manner consistent with those appearing in the area taxonomy of Figure 1.1. If a concept of a partial-area IS-A some concept (not necessarily the root) in another partial-area, then there is a *child-of* link pointing from the partial-area where the child concept resides to the partial-area where the parent resides. Consider *Abscess swab* mentioned earlier as an example. Since its two parents are in the partial-areas *Lesion sample* and *Swab*, respectively, the partial-area *Abscess swab* has two *child-of* links to those two partial-areas. If all the links (from partial-areas to partial-areas) were shown in a partial-area taxonomy, they might be too densely packed. To address this problem, the *child-of* links are further abstracted as follows: if all *child-of*'s in the same area (source area) have a common direction toward the partial-areas within another area (target area), then only one *child-of* is used to connect the source area to the target area. For example, in Figure 1.4(a), both partial-area *Abscess swab* and *Wound swab* point to the partial-areas *Lesion sample* in {*morphology*} and the partial-area *Swab* in {*procedure*}. So, only two *child-of*'s are used. This abstraction of *child-of* links is shown in Figure 1.4(b).

The partial-area taxonomy is designed in an effort to achieve hierarchical coherence in addition to the structural congruity of each of the areas. The network is a refinement of the area taxonomy. Figure 1.5 shows the (abridged) partial-area taxonomy of the Specimen hierarchy for the SNOMED July 2009 release. Each partial-area appears as a box inside its area node. In each partial-area node, the number in parentheses is the number of concepts it contains. For example, the area {*identity*} (second lower green box from left) is observed to have two partial-areas, *Device specimen* and *Specimen from* *patient*, of 19 and two concepts, respectively. All 19 concepts of the partial-area *Device specimen* represent specimens from devices. The partial-area taxonomy has a total of 361 partial-areas. An example *child-of* can be seen on the left side of the figure extending from partial-area *Effusion sample* to *Body substance sample*. Many *child-of*'s have been omitted.

Note that in some areas in Figure 1.5, the numbers of concepts in the partial-areas do not add up to the total number of concepts in the area in Figure 1.3. As a matter of fact, the area {*substance*} only contains 81 concepts, while the sum of concept numbers appearing in parentheses of its partial-areas is 136. This is due to overlaps among partial-areas, an issue that will be dealt with in detail in the following sections.

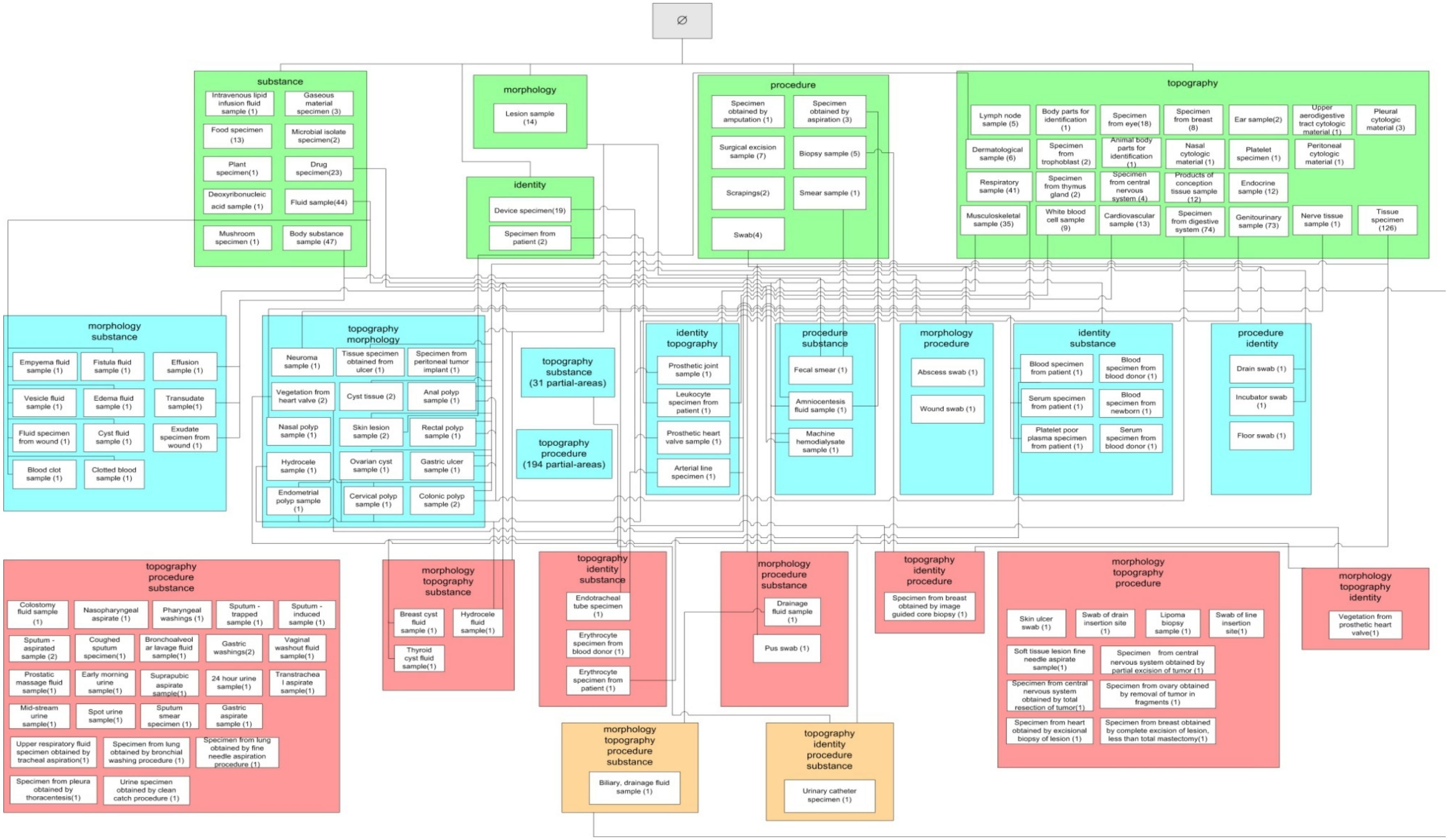


Figure 1.5 Partial-area taxonomy (abridged) for the Specimen hierarchy of the SNOMED July 2007 release.

1.2.4 SNOMED Auditing Techniques

Auditing is an important task of a terminology's life cycle [41]. Various auditing techniques have been applied to SNOMED. For example, comparative assessment of SNOMED's coverage and its completeness has been done [42-44]. The paper [45] proposed to use semantic methods to uncover concept classification errors. In [46], lexical information was used to detect classification omissions, and [47] focused on the inconsistent usage of "and" and "or" in SNOMED terms. In [48, 49], techniques were developed for discovering errors in concept hierarchies (e.g., cycles). The issue of balancing the problems of concept redundancy and ambiguity was addressed in [50]. In [51], a meta-level abstraction of the Semantic Network called a metaschema [52] was used to locate concepts having a high likelihood of errors. Ontological and linguistic techniques have been utilized to identify duplicates and redundancy [53, 54]. Bodenreider et al. [55, 56] analyzed how well SNOMED adheres to four basic ontological principles. The semantic completeness of SNOMED was assessed with a formal concept analysis (FCA)-based model in [57] and in [58] the impact of SNOMED revision was evaluated.

1.3 Dissertation Overview

This research builds upon and extends the previous work done by other members of the research group. This dissertation is an amalgamation of four papers that are organized as follows:

Chapter 2 [59] investigates an innovative auditing approach based on enhanced taxonomies, with Section 2.1 providing some background on how the enhanced taxonomy is formed to partition the overlapping concepts of partial-areas into disjoint sets. A thorough analysis of errors that are found as a result of auditing the overlapping concepts shows a need for enhancements to the partial-area taxonomy in order to capture a partition into disjoint sets having uniform semantics.

Chapter 3 [60] describes the converse abstraction network (CAN), dealing with one of the limitations of the taxonomy methodology, which relies heavily on the structure of the outgoing relationships of the concepts and is therefore not applicable to concepts with few or no relationships.

Chapter 4 [61] presents a preliminary study of the complementary use of the partial-area taxonomy to description logic (DL) for quality assurance purposes pertaining to SNOMED. Two kinds of errors are demonstrated in detail: missing IS-A relations and duplicate concepts. After correction, SNOMED is reclassified to ensure that no new inconsistencies were introduced.

Chapter 5 [62] brings up the idea of using taxonomies to measure the complexity and to track the evolution of a SNOMED hierarchy. The Specimen hierarchy is used as the test-bed. Two complexity measures are described: one is based on the relationship structure of the concepts and the other is based on the hierarchical arrangement of concepts with the same relationship structure.

Chapter 6 presents a preliminary exploration of the application of the partial-area taxonomy in a broader scope. Some methodologies that use the partial-area taxonomy to identify semantic-type assignment errors of SNOMED concepts in the UMLS are introduced. In particular, an application of an abstraction network based on the partial-area taxonomy to the Procedure hierarchy, which contains a large number of concepts and a rich set of relationships, is presented.

CHAPTER 2

AUDITING OVERLAPPING CONCEPTS OF SNOMED USING A REFINED HIERARCHICAL ABSTRACTION NETWORK

Auditors of a large terminology, such as SNOMED CT, face a daunting challenge. To aid them in their efforts, it is essential to devise techniques that can automatically identify concepts warranting special attention. “Complex” concepts, which by their very nature are more difficult to model, fall neatly into this category. A special kind of grouping, called a *partial-area*, is utilized in the characterization of complex concepts. In particular, the complex concepts that are the focus of this work are those appearing in intersections of multiple partial-areas and are thus referred to as *overlapping concepts*. In this research, an automatic methodology for identifying and partitioning the entire collection of overlapping concepts into disjoint, singly-rooted groups, that are more manageable to work with and comprehend, abstraction network for the overlapping concepts called a *disjoint partial-area taxonomy*. This new disjoint partial-area taxonomy offers a collection of semantically uniform partial-areas and is exploited herein as the basis for a novel auditing methodology. The review of the overlapping concepts is done in a top-down order within semantically uniform groups. These groups are themselves reviewed in a top-down order, which proceeds from the less complex to the more complex overlapping concepts. The results of applying the methodology to SNOMED’s Specimen hierarchy are presented. Hypotheses regarding error ratios for overlapping concepts and between different kinds of overlapping concepts are formulated. Two phases of auditing the Specimen hierarchy for two release of SNOMED are reported on. With the use of the double bootstrap and Fisher's exact test (two-tailed), the auditing of concepts and especially roots of overlapping partial-areas is shown to yield a statistically significant higher proportion of errors.

2.1 Disjoint Partial-Area Taxonomy

While areas serve to partition all the concepts of a SNOMED hierarchy, partial-areas do *not* do the same for the concept within an area. That is, a given concept may reside in more than one partial-area, a situation that occurs when the concept is a descendant of two or more roots. Such a concept is called an *overlapping concept*. An example of this can be seen in Figure 2.1, where the concept at the bottom *Dialysis fluid specimen* belongs to the partial-areas *Fluid sample* and *Drug specimen*. For another example, see Blood bag specimen from patient appearing twice, within two partial-areas, in {identity} (Figure 1.1).

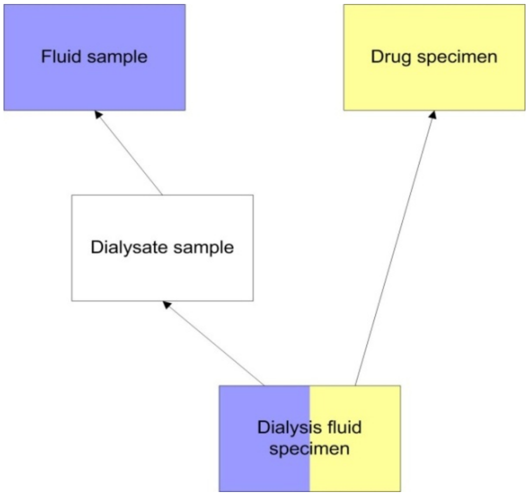


Figure 2.1 Overlapping concept *Dialysis fluid specimen*.

The presence of overlapping concepts somewhat degrades the categorization power of partial-areas. On the one hand, when looking at a specific partial-area, one can encounter concepts belonging solely to that partial-area and therefore elaborating the semantics of its root only. On the other hand, other concepts—the overlapping concepts—would belong to additional partial-areas at the same time and elaborate the semantics of multiple roots. The concept *Dialysis fluid specimen* from Figure 2.1 is both a fluid sample and a drug specimen, unlike its parent *Dialysate sample* which is only a kind of fluid sample. These situations cannot be determined at the level of the partial-area taxonomy. Moreover, overlapping concepts constitute knowledge convergence points within the hierarchy. As such, they warrant the designation “complex” and thus should be separated out from other concepts for the sake of auditing review.

In order to address these issues, the SABOC team has developed—in previous research [63]—an additional abstraction network, called the disjoint partial-area taxonomy, to properly model the overlapping portion partial-areas as nodes in their own right and therefore highlight the regions of complexity within the hierarchy. As it happens, the collection of overlapping concepts may represent a tangled hierarchy with many concepts exhibiting multi-parentage. The aim in formulating the disjoint partial-area taxonomy was impose some order on the overlapping concepts by partitioning them in such a way as to obtain a collection of concept groups satisfying single-rootedness. The details of the disjoint partial-area taxonomy can be found in [63]. In the following, the aspects relevant to the proposed auditing regimen are highlighted.

The basis for partitioning the overlapping concepts is the notion of overlapping root. Basically, such a concept is one that sits at the top of the overlapping concepts, with none of its parents themselves being overlapping concepts. In a recursive fashion, additional overlapping roots are identified below those at the very top of the overlapping section. As an illustration, the 15 overlapping roots of the area {*substance*} (2009) are shown as multi-colored boxes in Figure 2.2. The multi-coloring is used to indicate which area roots—appearing singly-colored at the top—the overlapping roots are descended from. For example, the overlapping root *Acellular blood (serum or plasma) specimen* is a descendant of *Blood specimen*, *Body substance sample*, and *Fluid sample*. The uncolored concepts are not roots of any kind.

Each overlapping root will be the root of its own newly formed concept group. The other concepts in a given group are, intuitively, those descendants of the root's (within the area) that are "between" it and other overlapping roots below. Collectively, these concept groups, called disjoint partial-areas (or d-partial-areas, for short) constitute a partition of the overlapping concepts of an area.

To obtain a complete partition of the area, all overlapping concepts are removed from the partial-areas in which they originally resided. This leaves the partial-areas with only non-overlapping concepts. For consistency, these modified partial-areas are referred to as d-partial-areas, too. Therefore, two varieties of d-partial-areas appear here: those having area roots and those having overlapping roots. For example, the d-partial-area rooted at *Body substance sample* consists of two concepts, the second being the non-overlapping concept *Stool specimen*. The d-partial-area rooted at the overlapping root *Acellular blood (serum or plasma) specimen* contains nine additional concepts, all of which are overlapping (see Figure 2.2).

The disjoint partial-area taxonomy is constructed from the d-partial-areas, which become nodes—within the area nodes—in the network. The labels of these nodes are the respective roots (either area or overlapping) of the d-partial-areas. The d-partial-area nodes are connected via *child-of*'s in a similar manner to those for areas and partial-areas. It should be noted that the d-partial-areas having area roots will appear at the top of the d-partial-area taxonomy; those with overlapping roots will be their descendants. The portion of the d-partial-area taxonomy for the area {*substance*} derived from the excerpt in Figure 2.2 can be seen in Figure 2.3.

As is shown in [63], there is an increase in the level of complexity of the roots of the d-partial-areas when proceeding in a top-down traversal of the d-partial-area taxonomy. Note that the other concepts in a given d-partial-area carry the same complexity as their root. This increasing complexity is taken into account in the auditing methodology introduced in this research.

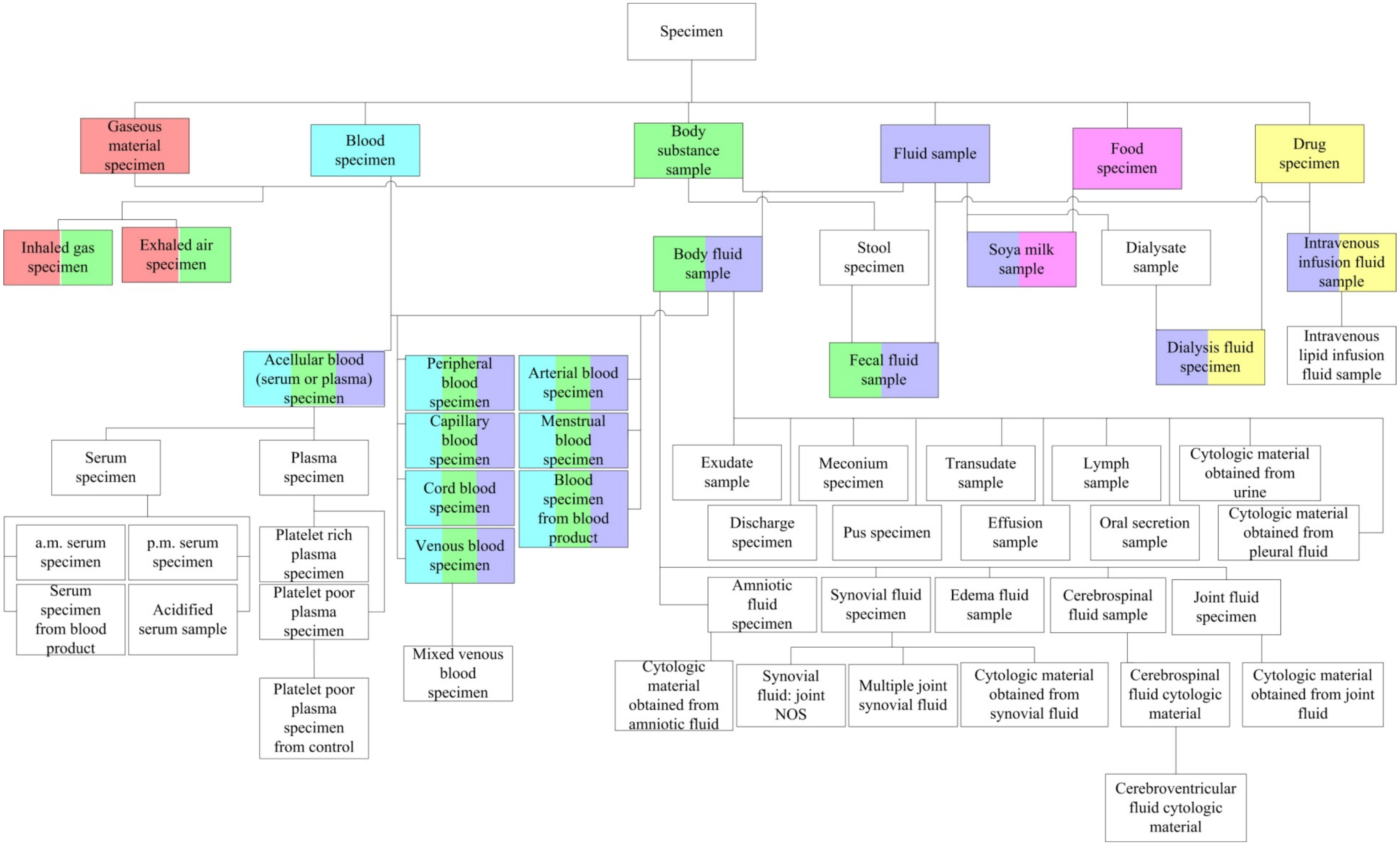


Figure 2.2 Overlapping roots in the Specimen hierarchy.

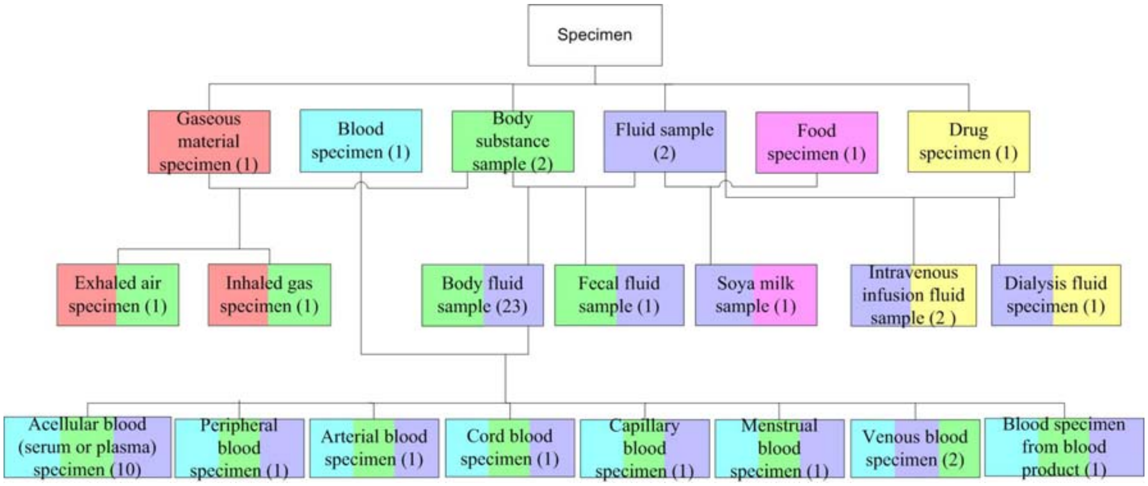


Figure 2.3 Portion of d-partial-area taxonomy corresponding to Figure 2.2.

2.2 Methods

As is discussed in the previous section, the overlapping concepts are complex concepts due to their multiple classifications with respect to the partial-area taxonomy and are thus targeted for auditing. Moreover, in [63], some overlapping concepts are seen to be more complex than others when moving down through the hierarchy. With these ideas in mind, the following auditing regimen is proposed that utilizes the paradigm of "group-based" auditing [2]. In the group-based approach applied to overlapping concepts, the concepts are reviewed in groups exhibiting semantic uniformity, that is, all the overlapping concepts of a d-partial-area are reviewed together with an eye toward the overlapping root which expresses the overarching semantics of the group. Furthermore, the concepts in the immediate neighborhood of the overlapping concepts (consisting of parents, children, siblings, and targets of relationships) are audited. This "neighborhood auditing" may help to uncover propagated errors, which might otherwise be missed if the review were limited to the overlapping concepts alone.

Since SNOMED is description-logic based [64], relationships are inherited by a child concept from its parent(s) along the IS-A hierarchy. Thus, an error such as an incorrect relationship will be inherited, too. Furthermore, even an error such as an omitted relationship may be "inherited" in the sense that if it is missing from the parent, it will probably be missing from the child (unless it is explicitly defined at the child).

As a consequence, it is preferred in an audit of a group of hierarchically related concepts that the review follows a top-down order. Following such an order may help in detecting more errors as well as in accelerating the review process. In particular, when a child is scrutinized, the auditor is already aware of any errors with the parents and is alert to their potential propagation. The topological sort [65] of a directed acyclic graph (DAG) – the structure exhibited by a SNOMED hierarchy – offers a traversal of concepts in a manner where each is processed only after all its parents have been processed. Because the d-partial-areas and their child-of relationships also constitute a DAG [63], the disjoint partial-area taxonomy enables the utilization of the topological sort order at two different levels: the d-partial-area level and the concept level, with the latter nested in the former.

The following describes the auditing methodology for overlapping concepts based on the disjoint partial-area taxonomy. It should be noted that overlapping roots come in two varieties: *base* and *derived*. The details can be found in [63]. The important distinction between the two in this context is that the base overlapping roots occur toward the top of the concept hierarchy and are above all the derived overlapping roots. Also note that some d-partial-areas do not have any overlapping concepts at all. They are the ones at the very top of the disjoint partial-area taxonomy that were residually left over after the lower-level d-partial-areas – containing overlapping concepts – were removed from their original partial-areas. For example, the top d-partial-area *Drug specimen* (1), comprising a single, non-overlapping concept, was left over as a result of extracting the d-partial-areas *Intravenous infusion fluid sample* (2) and *Dialysis fluid specimen* (1) (see Figure 6) from the partial-area also named "*Drug specimen*" that contained a total of four concepts. Those upper-level d-partial-areas are not considered in the auditing methodology.

1. **Taxonomy level**: The d-partial-areas are processed in topological sort order starting with those having base overlapping roots. The processing proceeds through their children, grandchildren, etc., down to the very bottom of the disjoint partial-area taxonomy. As discussed in [63], the lower d-partial-areas are rooted at more complex overlapping concepts.
2. **Concept level**: On arrival at a particular d-partial-area in (1), all its constituent concepts are reviewed in a topological sort order starting with its unique root and processing downwards. The concepts are presented to the auditor in an indented hierarchy (textual) former for inspection (see, e.g. Figure 1.1). The indented display neatly supports the top-down processing where each concept is reviewed only after all its respective parents are reviewed.

Please note that the topological sort order leaves degrees of freedom with regards to the order with which the nodes of the graph are visited – and reviewed. For example, in a level-by-level traversal, all nodes on a given level are processed before any node on the next level. Another choice is a "preorder traversal," where the processing proceeds from a parent node to its children and even its grandchildren, assuming all their parents were already processed at that point. For the effectiveness of the auditing regimen, the preorder traversal is recommended. In this way, the scrutiny of a child follows that of the parent as quickly as possible, allowing an auditor to more readily retain knowledge of errors discovered at the parent and potentially propagating to the child.

To illustrate the Taxonomy level, the review will begin with the bicolored d-partial-areas in Figure 2.3, including *Exhaled air specimen*, *Inhaled air specimen*, etc. Once the review reaches *Body fluid sample*, the only bicolored d-partial-area with children, it proceeds to the bottom level containing eight tricolored d-partial-areas, i.e., *Acellular blood (serum or plasma) specimen*, *Peripheral blood specimen*, and so on. When all child d-partial-areas of *Body fluid sample* have been audited, the processing continues with the rest of the bicolored d-partial-areas, e.g., *Dialysis fluid specimen*. Again, the d-partial-areas of one color in Figure 2.3 do not have overlapping concepts and are therefore not part of the auditing regimen.

Within the d-partial-area *Body fluid sample*, the Concept level processing would begin with the root *Body fluid sample* and then proceed to its 22 children, including *Exudate sample* and *Discharge specimen* (Figure 2.2). When a concept with children is encountered, the children are processed immediately after the parent to support the auditor in detecting error propagation from parent to child. For example, *Amniotic fluid specimen* is followed by its child *Cytologic fluid specimen* obtained from amniotic fluid. An example of a propagation of an error that is easily detectable when reviewing a d-partial-area can be seen with the concept *Synovial fluid specimen* in the d-partial-area *Body fluid sample* (Figure 2.2). A missing topography relationship is detected with the target *Articular space* in the Body Structure hierarchy. The same missing relationship is detected for its three children: *Multiple joint synovial fluid*, *Cytologic material obtained from synovial fluid*, and *Synovial fluid joint NOS*. Arriving later at the d-partial-area *Acellular blood (serum or plasma) specimen*, the root would be examined first. Note that the root's overlapping parent *Body fluid sample* would already have been examined according to the Taxonomy level ordering. The review of its child *Serum specimen* and its four children would follow. Only after that would the review of the sibling *Plasma Specimen* and its three descendants occur (see Figure 2.2).

To test the methodology, all the overlapping concepts of SNOMED's Specimen hierarchy are audited. All areas of that hierarchy are considered.

Two hypotheses are desired to be investigated in regard to this study. The first distinguishes between overlapping concepts and non-overlapping concepts. The second distinguishes between overlapping roots of d-partial-areas and other overlapping concepts.

Hypothesis 1: Concepts residing in d-partial-areas having overlapping roots (i.e., overlapping concepts) are more likely to have errors than concepts residing in d-partial-areas containing no overlapping concepts.

Hypothesis 2: Overlapping roots of d-partial-areas are more likely to have errors than non-root overlapping concepts.

The first hypothesis asserts that these more complex concepts indeed exhibit a higher number of errors. The second hypothesis refers to the more significant overlapping concepts as the overlapping roots, where the convergence of multiple inheritance paths occurs and where higher concentrations of errors are expected.

As a basis for comparison a control sample comprising concepts gleaned from partial-areas having no overlaps whatsoever are also audited. Both kinds of concepts are audited by the same auditors. To compare overlapping concepts with those in the control sample, the proportion of erroneous concepts is checked. The d-partial-area is used as the unit of analysis, and across levels (because of the small number of concepts at Level 2) are aggregated. Both hypotheses are tested for Phases 1 and 2 of the auditing on the two releases of SNOMED, two years apart. The double bootstrap [66] is employed and Fisher's exact test two tailed [67] to calculate the statistical significance of the difference of the proportions, for Hypothesis 1 and 2, respectively.

2.3 Results

Two phases of results obtained with respect to two releases of SNOMED are reported. Phase 1 for the July 2007 release and Phase 2 for July 2009. In Phase 2, the methodology described in the previous section is utilized and based on the disjoint partial-area taxonomy. During Phase 1, the methodology was not yet developed and therefore and exhaustive audit of all overlapping concepts was carried out without regard to any structural configuration or ordering. A preliminary report with some results of Phase 1 appeared in [68].

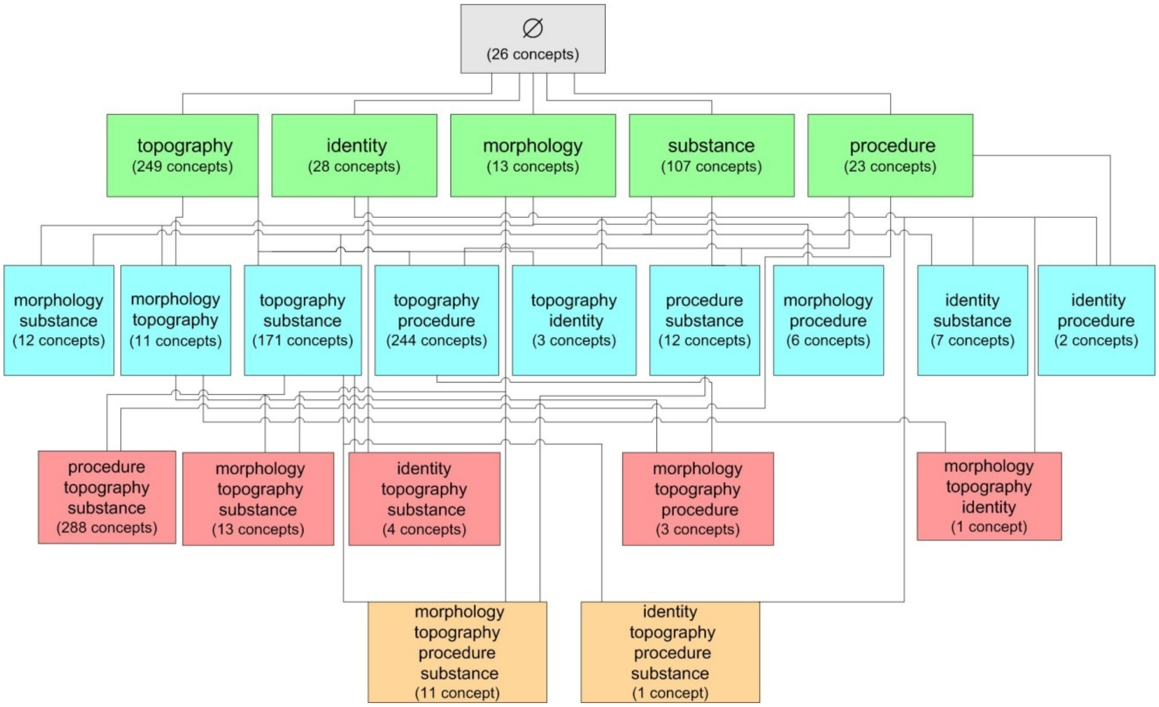


Figure 2.4 Area taxonomy for Specimen hierarchy (2009).

The results of Phase 1 (see Figure 1.3 and Figure 1.5) were reviewed by Dr. Kent A. Spackman (currently the Chief Terminologist of IHTSDO [1]) and those approved by him were submitted to CAP [69] for consideration and incorporation into the Specimen hierarchy. As a result, there were many changes in the overlapping concepts of this hierarchy as reflected in SNOMED’s July 2009 release. The area taxonomy and the partial-area taxonomy for the July 2009 release appear in Figure 2.4 and 2.5, respectively. A comparison of the area taxonomies of 2007 (Figure 1 in [63]) and 2009 (Figure 2.4) exposes many differences in the Specimen hierarchy. For example, the total number of concepts with one relationship – which is equal to the sum of the sizes of the (green) areas on Level 1 – went down from 468 to 420. At the same time, the area {*substance*} grew from 81 to 107 concepts. The number of areas with three relationships went down from seven to five with the loss of the two areas

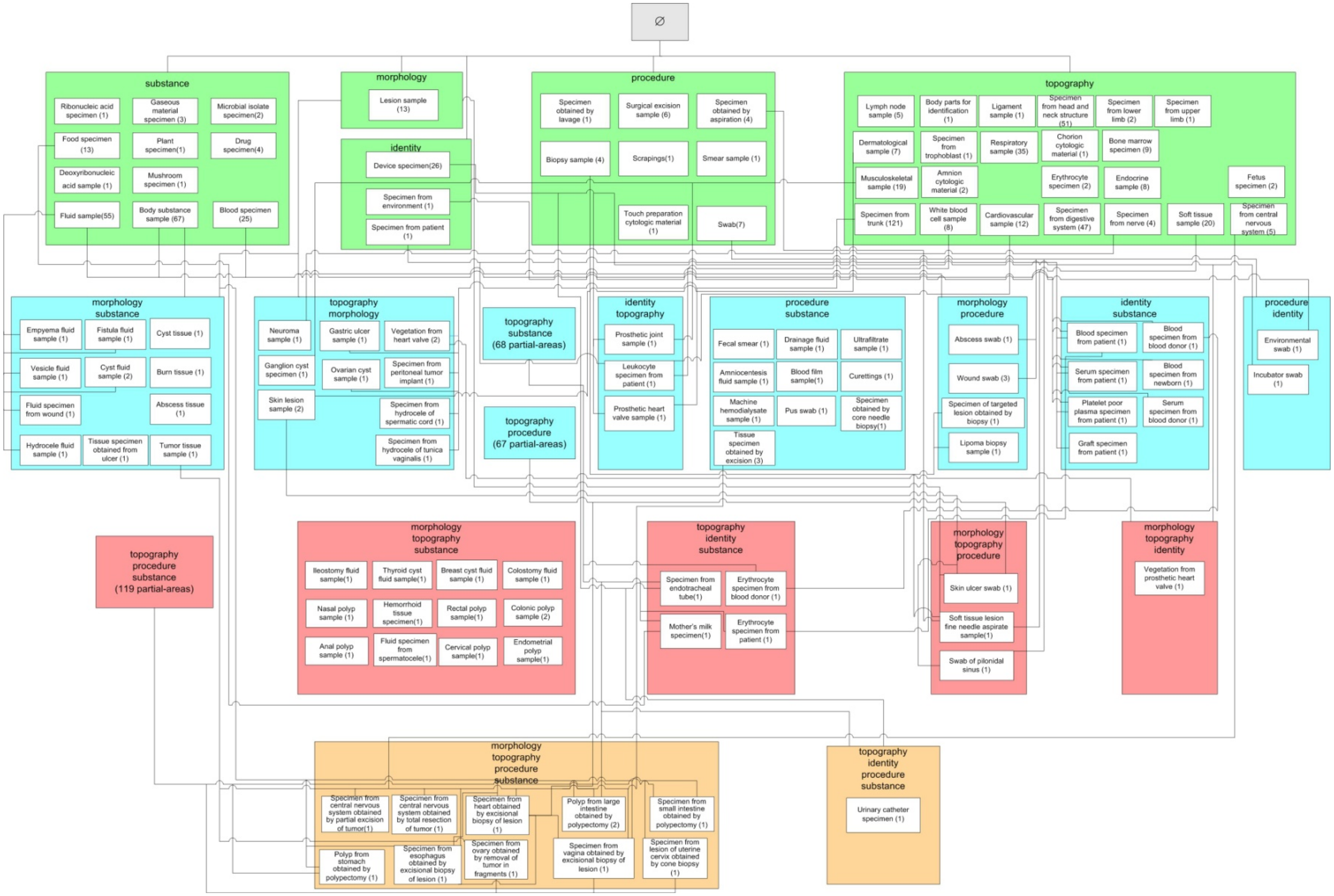


Figure 2.5 Specimen partial-area taxonomy (2009).

{*morphology*, *procedure*, *substance*} and {*topography*, *identity*, *procedure*}. On the other hand, the area {*procedure*, *topography*, *substance*} grew from 26 concepts in 2007 to 288 concepts in 2009.

Similarly, comparing the partial-area taxonomies for 2007 and 2009 reveals many differences. For example, the area {*substance*} changed from having ten to 11 partial-areas. But that small numerical change is misleading, as one can guess, considering the 32% increase in the size of the area. Only six partial-areas did not change. A new partial-area is *Blood specimen* with 25 concepts. At the same time, *Drug specimen* shrank from 23 to four concepts, mainly due to the removal of blood specimen concepts. *Body substance sample* expanded from 47 to 67 concepts, while *Fluid sample* grew from 44 to 55 concepts. Such large changes on the partial-area level seem to indicate an increase in the overlap size when compared to the overall increase of 26 concepts observed on the area level. As another example, the area {*morphology*, *topography*, *substance*} went from having three partial-areas to 12. The area {*morphology*, *topography*, *procedure*, *substance*} grew from one to ten.

The number of overlapping concepts increased by 48 from 162 to 210 (30%). Clearly, the landscape of the overlapping portions of partial-areas changed meaningfully from the time of the July 2007 release. For example, as was predicted above, in the area {*substance*}, there were 35 overlapping concepts in nine d-partial-areas in 2007 (Figure 9 in [63]), but 48 overlapping concepts in 15 d-partial-areas in 2009 (Figure 2.3).

The SABOC team decided that this change warranted the Phase 2 auditing involving the July 2009 release's overlapping concepts. This decision was also motivated by the opportunity to apply the new auditing methodology introduced in Section 2.1. The expectation was that the new methodology employing a detailed order of review would expose errors missed during Phase 1. The results of Phase 2 were reviewed by Dr. James T. Case of the SNOMED US National Release Center (NRC) at the NLM for possible inclusion in the US extension of SNOMED. Any changes approved by him are transferred to the IHTSDO for review and potential inclusion in SNOMED’s international release. The results of the reviews conducted by Dr. Spackman for the 2007 release and Dr. Case for the 2009 release serve in this study as validations of the errors uncovered by the auditors. Only those approved errors are used in the analysis that follows.

2.3.1 Phase 1: Auditing of July 2007 SNOMED

The July 2007 release of the Specimen hierarchy consists of 1,056 concepts, of which 162 are overlapping. For its partial-area taxonomy, see Figure 2 in [63]. Most of the overlapping concepts reside in Level 1 areas, i.e., those having one relationship. In fact, roughly one third (155 out of 468) of the Level 1 concepts are overlapping. And these are found primarily in the area {*topography*} and {*substance*}. The d-partial-areas of {*substance*} and {*topography*} can be seen in Figure 9 and 10, respectively, in [63]. Overlapping concepts also appear in the partial-areas of areas with two relationships but in far fewer numbers. In fact, there are only seven of them. Six are in {*topography*, *procedure*}, and the other is in {*topography*, *morphology*}.

For Phase 1, two domain-expert are enrolled (GE and JX), each of whom has training in medicine as well as training and experience in medical terminologies. The overlapping concepts were reviewed individually by both of them. Their review did not follow the current auditing methodology – newly presented herein – but instead involved an examination of all such concepts without regard to any specific order [68]. After that, the two together reviewed concepts for which their individual reports differed, and discussed the discrepancies until a consensus was reached. A consensus report was given to Dr. Spackman for further review, and only his accepted results are reported in this section for Phase 1. Note that the current new a posteriori analysis of the results and the assessment of the hypotheses are based on the new disjoint partial-area taxonomy presented in [63].

Table 2.1 presents the results of auditing the 35 overlapping concepts (see Figure 8 in [63]) distributed across nine d-partial-areas in the area {*substance*} (Figure 9 in [63]). For each d-partial-area, the following are listed: number of overlapping concepts V, number of erroneous overlapping concepts V_{err}, the number of errors E_{root} exhibited by the overlapping root, and the total number of errors E for all overlapping concepts.

Table 2.1 Auditing Results for Overlapping Concepts of {*substance*} Arranged by D-partial-area

D-partial-area	**V**	**V_{err}**	**E_{root}**	**E**
Exhaled air specimen	1	0	0	0
Inhaled gas specimen	1	0	0	0
Fecal fluid sample	1	0	0	0
Acellular blood (serum or plasma) specimen	1	1	1	1
Serum specimen from blood product	1	1	3	3
Serum specimen	2	0	0	0
Plasma specimen	4	1	1	1
Body fluid sample	11	3	17	19
Blood specimen	13	5	2	7
Total:	**35**	**11**	**24**	**31**

V = # overlapping concepts; V_{err} = # erroneous overlapping concepts;
E_{root} = # errors at the overlapping root; E = total # errors at overlapping concepts

For example, the largest d-partial-area *Blood specimen* has 13 concepts, of which five were found to be in error. The root *Blood specimen* had two errors, and overall the d-partial-area's concepts had seven. For this d-partial-area, 50% (six out of 12) of the non-root overlapping concepts are erroneous, while the root itself exhibits two errors. The result, for one example of a d-partial-area, gives support to Hypothesis 2.

Table 2.2 Auditing Results for Overlapping Concepts by Area

Area	**C**	**V**	**D**	**V_{err}**	**E**	**E/ V_{err}**	**E/V**	**D_{err}**	**E_{root}**	**E_{root}/ D_{err}**	**D_{err} /D**	**(V_{err}-D_{err}) /(V-D)**
substance	81	35	9	11	31	2.8	0.89	5	24	4.8	56%	23%
topography	333	116	52	71	110	1.6	0.95	39	62	1.59	75%	50%
procedure	20	3	3	3	9	3.0	3.0	2	9	4.5	66%	N/A
identity	20	1	1	0	0	N/A	0	0	0	N/A	0%	N/A
topog., proc.	380	6	6	4	9	2.3	1.5	4	9	2.3	66%	N/A
topog., morph.	18	1	1	0	0	N/A	0	0	0	N/A	0%	N/A
Total:	**852**	**162**	**72**	**89**	**159**	**1.8**	**0.93**	**50**	**104**	**2.1**	**69%**	**43%**

C = # concepts; V = # overlapping concepts; D = # overlapping roots;
V_{err} = # erroneous overlapping concepts; E = total # errors at overlapping concepts;
D_{err} = # erroneous overlapping roots; E_{root} = # errors at the overlapping roots; N/A = Not applicable

The auditing results for all overlapping concepts are listed by area Table 2.2. For each area, its total number of concepts C, number of overlapping concepts V, number of overlapping roots D, number of erroneous overlapping concepts V_{err}, total number of errors E for the overlapping concepts, number of erroneous overlapping roots D_{err}, number of errors E_{root} exhibited by the set of overlapping roots, and a number of relevant ratios are shown. For example, {*substance*} has 81 concepts, of which 35 are overlapping. Eleven (31%) of the latter were found to have a total of 31 errors or an average of 2.8 per erroneous concept, as detailed in Table 2.2. The ratio of the total number of errors at the overlapping concepts to the number of overlapping concepts is 0.89. Of the nine overlapping roots, five (56%) were found to be in error – with a combined 24 errors among them (or 4.8 errors per erroneous root). But only 23% (= (11-5)/(35-9)) of the non-root overlapping concepts had errors. Let us note that for some areas (e.g., {*procedure*}), the ratio in the last column is not applicable (undefined) since singletons (i.e., d-partial-areas containing just one concept) have no non-root overlapping concepts. Other ratios may not be applicable due to a lack of errors. Nevertheless, the total ratios at the bottom of the table are defined across all the areas with overlapping concepts.

Table 2.3 Results of Auditing Intersections Involving partial-area *Tissue specimen*

Second Partial-Area	**C**	**V**	**V_{err}**	**V_{err} / V (%)**
Specimen from eye	18	12	8	67
Ear sample	2	1	0	0
Specimen from breast	8	4	2	50
Cardiovascular sample	13	3	1	33
Products of conception tissue sample	12	1	1	100
Genitourinary sample	73	20	17	85
Dermatological sample	6	2	0	0
Specimen from digestive system	74	29	18	62
Musculoskeletal sample	35	22	15	68
Respiratory sample	41	6	5	83
Endocrine sample	12	3	0	0
Specimen from central nervous system	4	1	0	0
Spec. from thymus gland	2	1	0	0
Specimen from trophoblast	2	1	0	0

Most overlapping concepts in {*topography*} are found in intersections of partial-areas involving *Tissue specimen* containing 126 concepts. These results have been tabulated separately in Table 2.3. For example, the partial-area *Specimen from eye* has 18 concepts. Its intersection with *Tissue specimen* has 12 of them. Eight of those are in error.

The control sample was gleaned from partial-areas from partial-areas that had no intersections whatsoever with other partial-areas and from d-partial-areas having no overlapping concepts (i.e., those left over after the removal of the d-partial-areas with overlapping concepts from a partial-area; see, e.g., the six d-partial-areas at Level 1 of Figure 2.3). Furthermore, only partial-areas that contained more than one concept are used. The reason for the last requirement is that, as alluded to, partial-areas of one concept are already known to be error-prone [41, 70]. Thus, they do not make for a proper control sample.

A control sample of 78 concepts is used from Level 1, half of its overlapping concepts (155). From Level 2, seven concepts are gathered for the control sample, an equal number to the overlapping concepts. Hence, there are 155+7=162 overlapping concepts, and the control sample has 78+7 = 85 concepts. Since the purpose was to audit overlapping concepts, a smaller control sample is used that was large enough to support statistical significance for the result presented below.

Table 2.4 Auditing Results for Overlapping Concepts vs. Control Sample (Phase 1)

	C	E	E / C	C_{err}	$C_{err}/C(\%)$	E/C_{err}
Overlapping	162	158	0.98	89	55	1.8
Control Sample	85	31	0.36	25	29	1.2

Table 2.4 gives the results of the auditing carried out on these two groups of concepts. C denotes the number of concepts, E (Column 3) denotes the total number of errors, and C_{err} is the number of erroneous concepts (Column 5) – with a given concept potentially having more than one error. The average erroneous – concept rate among the overlapping concepts was 55%, and among the control sample it was 29% (Column 6). The difference was significant (using the double bootstrap [71]) at the 0.05 level, supporting Hypothesis 1. Let us point out that there was nearly one error (0.98) on average per overlapping concept as compared to 0.36 on average within the control sample (Column 4). Moreover, erroneous concepts in the overlapping group had 1.8 errors on average (last column) versus 1.2 errors on average for the control sample, showing further difference between the two.

In examining the auditing results, overlapping roots are found to be more error-prone than other overlapping concepts. For example, in {*procedure*} and {*topography*, *procedure*}, all errors are found in overlapping roots. As shown in Table 2.2, in the area {*substance*}, five out of nine roots (55%) versus six (= 11-5) out of 26 (=35-9) non-root overlapping concepts (23%) were found to be erroneous. To assess Hypothesis 2, the data from Table 2.2 are used for the entire collection of overlapping concepts. The percentage of erroneous concepts for overlapping roots is 69% (=50/72). The percentage of erroneous concepts in the set of non-root overlapping concepts is 43% (=(89-50)/(162-72)). The difference in the percentages of erroneous concepts between the overlapping roots (69%) and the non-root overlapping concepts (43%) is statistically significant (Fisher's exact test two-tailed [67], p-value = 0.0014), supporting Hypothesis 2.

2.3.2 Phase 2: Auditing of July 2009 SNOMED

For Phase 2, three domain-experts (GE, JX, and YC) are enrolled, each of whom has training in medicine as well as training and experience in medical terminologies. The overlapping concepts were reviewed individually by each of the auditors, according to the methodology described above. That is, the concepts of each d-partial-area's root. Furthermore, if one d-partial-area is the child of another, the review of the concepts of the parent precedes the review of the concepts of the child. The experts reviewed the overlapping concepts for eight types of errors, enumerated on a given form. Their findings were anonymized and summarized. The three experts were requested to review the summarized report, and they marked whether they agreed or disagreed with the errors listed.

Table 2.5 Sample of Error Types of Overlapping Concepts for July 2009 Release

Concept	Partial-areas	Error Type(s)	Correction(s)
Serum specimen from blood product	Blood specimen / Fluid sample/Body substance sample	Missing parent	Add parent: Blood specimen from blood product
Dentin specimen	Specimen from digestive system/Specimen from head and neck structure	Incorrect Parent: Oral cavity sample	Correct parent: Specimen from tooth
a.m. serum specimen	Blood specimen/Fluid sample(specimen)/Body substance sample	Missing relationship	Add relationship: TIME_ASPECT with the value of – am-ante meridiem
Specimen from tooth	Specimen from digestive system/Specimen from head and neck structure	Incorrect relationship target: Oral cavity structure	Refine with: Tooth structure
Specimen obtained by fine needle aspiration procedure	Specimen obtained by aspiration/Biopsy sample	Missing child	Add children: *Breast fine needle aspirate sample; *Soft tissue lesion fine needle aspirate sample; *Specimen from heart obtained by fine needle aspiration procedure; *Specimen from thymus gland obtained by fine needle aspiration biopsy
Tissue specimen from placenta	Tissue specimen from genital system/Products of conception tissue sample	Other error type: missing ancestor “Soft tissue sample”	Create a proper concept to parent it in the “Soft tissue sample” tree.

In the Phase 2 review, a better agreement regarding the combined reported results is tried to be achieved. One expert might have overlooked an error discovered by another, and may have agreed with it, once the potential error was reported. The level of agreement improved after the second-stage review. All overlapping concepts are reported as potential errors to the SNOMED United States NRC having at least one auditor reporting an error for them. The report was reviewed by Dr. Case (who works at the NRC). Only errors confirmed by him are considered in the results presented in the following.

A sample of different types of errors agreed upon by all three auditors is listed in Table 2.5. For example, it was agreed that *Serum specimen from blood product* is missing a parent *Blood specimen from blood product* that should be added. Table 2.6 summarizes the number of occurrences for each type of error found in the overlapping concepts of the July 2009 release reported to the NRC. Missing parents, for example, were found for 23 concepts.

Table 2.6 Distribution of Types of Errors in the Second Phase of Auditing Overlapping Concepts

Error Type	# Concepts
Missing parent	23
Incorrect parent	22
Missing child	6
Incorrect child	2
Missing relationship	55
Incorrect relationship target	2
Other error type	6

The auditing results for Phase 2 are listed by area in Table 2.7, in the same format used in Table 2.2 for Phase 1. In this case, for example, {*topography*} has 249 concepts, with 110 of them being overlapping. Fifty-two out of the 110 (47%) were found to have a total of 57 errors or an average of 1.10 per erroneous concept. The ratio of the total number of errors to the number of overlapping concepts is 0.52. Twenty of the 37 overlapping roots (54%) were found to be in error – with a combined 22 errors among them (or 1.10 errors per root). Finally, 44% (=(52-20)/(110-37)) of the non-root overlapping concepts had errors.

For the entire set of overlapping concepts summarized in the bottom row of Table 2.7, 127 out of 210 (60%) were found to be erroneous. This result is applicable in assessing Hypothesis 1 (as shown in Table 2.8).

Table 2.7 Phase 2 Auditing Results for Overlapping Concepts by Area

Area	**C**	**V**	**D**	**V_{err}**	**E**	**E/ V_{err}**	**E/ V**	**D_{err}**	**E_{root}**	**E_{root} /D_{err}**	**D_{err} /D (%)**	**(V_{err}-D_{err}) /(V-D)**
substance	107	48	15	28	36	1.29	0.75	8	11	1.38	53	61%
topography	249	110	37	52	57	1.10	0.52	20	22	1.10	54	44%
procedure	23	2	1	1	1	1.00	0.50	1	1	1.00	100	0%
topog., proc.	244	29	16	28	38	1.36	1.31	15	19	1.27	94	100%
topog., subst.	171	5	4	3	4	1.33	0.80	3	4	1.33	75	0%
subst., topog., proc.	288	16	14	15	25	1.67	1.56	14	23	1.64	100	50%
Total:	1,082	210	87	127	161	1.27	0.77	61	80	1.30	70	54%

C = #concepts; V=#overlapping concepts; D=#overlapping roots;
V_{err} = #erroneous overlapping concepts; E=total #errors;
D_{err} = # erroneous overlapping roots; E_{root} = #errors at the roots;

The control sample for Phase 2 was taken strictly from partial-areas and d-partial-areas that had no intersections whatsoever. As with Phase 1, only partial-areas that contained more than one concept are used. The sample consisted of 111 concepts from the same areas as the overlapping concepts. And as in Phase 1, the number of sample concepts taken from areas with small numbers (i.e., 2 – 16) of overlapping concepts was about the same as the number of overlapping concepts taken from those areas. The sample concepts numbered about half the overlapping concepts for areas with larger numbers of overlapping concepts. As with Phase 1, the purpose was to audit overlapping concepts, and a smaller control sample is used that was nevertheless big enough to support statistical significance of the result.

Like Table 2.4, Table 2.8 juxtaposes the results of auditing the overlapping concepts and those in the control sample. The average erroneous-concept rate among the overlapping concepts was 60%, versus 31% for the control sample (Column 6). The difference was significant at the 0.05 level, supporting Hypothesis 1. Let us note that there were 0.77 errors on average per overlapping concept as compared to 0.32 on average within the control sample (Column 4). Erroneous concepts in the overlapping group had 1.27 errors on average (last column) versus 1.03 errors on average for the control sample, showing further difference between the two samples.

Table 2.8 Auditing Results for Overlapping Concepts vs. Control Sample (Phase 2)

	C	E	E/C	C_{err}	C_{err}/C(%)	E/C_{err}
Overlapping	210	161	0.77	127	60	1.27
Control Sample	111	36	0.32	35	32	1.03

For the assessment of Hypothesis 2, the results obtained for all overlapping concepts are used, reflected in the bottom row of Table 2.7. Among the 87 overlapping roots, 61 (70%) were erroneous, while for the 123 (=210 - 87) non-root overlapping concepts, 66 (=210-87 or 54%) were found to be in error. The difference in the percentages of erroneous concepts between the overlapping roots (70%) and the non-root overlapping concepts (54%) is statistically significant (Fisher's exact test two-tailed, p-value = 0.0217).

2.4 Discussion

2.4.1 Auditing Theme: Complex Concepts

This study is motivated by a general theme that more "complex" concepts tend to have more errors than simpler concepts. The theme of being more complex may manifest itself in a variety of ways. One manifestation of this theme for partial-areas was the group of concepts residing in "strict inheritance" partial-areas [70]. In the context of the present work, this theme appears twice: the first time in identifying overlapping concepts as more complex than non-overlapping concepts due to their elaborating the multiple semantics of the multiple partial-areas they belong to; the second in the distinction between overlapping roots and non-root overlapping concepts. The reason for the higher complexity of overlapping roots stems from their being at the junction points where multiple hierarchical paths from ancestors converge. Each such path contributes a portion of a diverse collection of inherited knowledge at the overlapping root. Hypothesis 1 addresses the first appearance. Hypothesis 2 pertains to the second.

As was also shown in [70] with regards to strict inheritance partial-areas, the results of the study confirm the auditing theme that complex concepts have relatively more errors. In view of the fact that modeling complex concepts is more challenging than modeling simpler concepts, it is not really surprising to find more errors in the former. The research challenge is to discover various characterizations of “complex” concepts. In particular, it is fruitful to identify structural characterizations that can be computed automatically, as in the current study and in [70]. The higher error rate shown here and in [70] will help achieve higher productivity from quality-assurance personnel in their review of such concepts. It is suggested that the design of partial area taxonomies and the auditing of the complex concepts discussed here and in [70] should become integral parts of the design cycle for terminologies such as SNOMED and the NCIt [41]. Such techniques will also help interface terminologies such as Kaiser-Permanente’s CMT [72] or the VA’s ERT [73], which were derived initially from SNOMED and were enhanced with local vocabulary as well as integrated parts of other terminologies. It is a research challenge to identify more manifestation of complex concepts using taxonomies or other structural techniques for SNOMED and similar terminologies.

One may wonder why there are more errors in overlapping roots than there are in other overlapping concepts (as stated in Hypothesis 2), in spite of the expectation that this methodology will expose error propagation from parents to children, which implies that errors at an overlapping root would be "inherited" by the other concepts in its d-partial-area. One should realize that indeed missing or incorrect relationship errors are "inherited," but that is not true of other errors, e.g., an incorrect parent. Furthermore, many d-partial-areas have just a single concept (which serves as the respective root), with no children below to inherit the errors. Hence, this methodology is designed to expose the cross-generational error propagation to the extent that it exists.

2.4.2 Repeated Application of an Auditing Methodology

In previous research [2, 70], various methodologies for auditing a SNOMED hierarchy are presented. A question to consider is whether there is a reason to reapply the same auditing technique to the hierarchy obtained following corrections derived from the earlier auditing phase that used the same technique. Should it be assumed that not all errors were found and corrected? In the context of this research, the question was: should the overlapping concepts be audited again following the first phase reported in [68]? Furthermore, how many times should the same technique be applied? Another way to phrase this last question is: how do we identify the convergence of the auditing process?

We had several reasons to re-audit the overlapping concepts. First, in Phase 1, we just audited the set of all overlapping concepts without utilizing any structure among them. In this paper, we introduced the new "group auditing" methodology of overlapping concepts where d-partial-areas were utilized as the grouping unit following the new framework described in [63]. Furthermore, the new methodology employs a top-down ordering within each d-partial-area and among various d-partial-areas.

Another reason for repeating the auditing on the overlapping concepts is the large increase in their numbers and the number of d-partial-areas. For example, see Figure 2.3 for the d-partial-areas in the area {*substance*} in comparison to the corresponding Figure 9 that appeared in [63]. In Figure 9 of [63], only four d-partial-areas without overlapping concepts are seen at the first level and nine d-partial-areas comprising overlapping concepts. In Figure 2.3, showing the overlapping concepts of {*substance*} in 2009, there are six top d-partial-areas without overlapping concepts and 15 d-partial-areas with overlapping concepts. Moreover, when one reviews the details of the two figures, many internal changes can be seen. For example, the d-partial-area *Body fluid sample* had 11 concepts in 2007 and 23 in 2009. *Blood specimen* had 13 overlapping concepts in Level 3 originally, and in 2009 it is a top d-partial-area of one concept only. It has eight child d-partial-areas containing 18 overlapping concepts on Level 3, which are shared jointly by the parent d-partial-area *Body fluid sample* (see Figure 2.3). The latter was a parent of *Blood specimen* in Figure 9 of [63]. Obviously, such changes reflect an entire remodeling of many overlapping concepts.

When realizing the extent of the changes, it was possible that new errors were introduced and that the new d-partial-taxonomy would lead to exposure of errors not reported in the review of the 2007 release. The results shown in Table 2.7 justify the decision for the second auditing phase. While a meaningful amount of errors are expected to be found in Phase 2, it is surprising by their magnitude. Both the percentages of the erroneous concepts among overlapping concepts (60% vs. 55%) and among overlapping roots (70% vs. 69%) were little changed in spite of this being a second round of auditing. Part of the explanation may be the improved methodology employed in this study. Another reason may be the large increase in the number of overlapping concepts (from 162 to 210). A further factor might be that in practice the proper modeling of these complex concepts demands more than one iteration.

On the other hand, the ratio of errors per erroneous concept was reduced (0.93 to 0.77) for all overlapping concepts, as was the ratio for erroneous overlapping roots (2.1 to 1.3). Hence, while the percentage of erroneous concepts persisted, the average number of errors fell. That is, fewer concepts with multiple errors are found. This last observation seems in line with the speculation above that multiple iterations are required for the proper modeling of complex concepts.

One could certainly question the expectation of the need for an additional phase of auditing after all corrections from the overlapping concept regimen have been implemented. That is particularly true when the corrections have made their way into SNOMED's international release following the report of Dr. Case (at the NRC) to IHTSDO. To better understand the phenomenon of finding more errors in a subsequent phase of auditing overlapping concepts mentioned above, one needs to keep in mind the restructuring undergone by d-partial-areas due to the discovered errors. For example, in the description of the methodology in Section 2.1, a concept *Synovial fluid specimen* in the d-partial-area *Body fluid sample* are mentioned, which together with its children is missing the relationship *topography* to *Articular space*. But reviewing the complete audit report for the overlapping concepts in {*substance*}, one may realize that the same concept was found to have an incorrect parent, *Body fluid sample*, which was replaced by *Joint fluid specimen*. This latter concept was independently found to be missing the same *topography* relationship, as was its child *Cytologic material obtained from joint fluid*. Furthermore, another concept *Synovial fluid cells* in the area {*topography*} was also made a child of *Synovial fluid specimen* instead of *Synovial sample*. What is seen is that a movement of many concepts into the d-partial-area rooted at *Joint fluid specimen*, which before had only one child. Moreover, this d-partial-area would move from the area {*substance*} to the area {*substance*, *topography*} due to the additional *topography* relationship. When all these corrections are incorporated into a future release of SNOMED, the d-partial-area taxonomy will convey the refined modeling of all joint fluid specimen concepts, contributing to better overall comprehension. However, this new modeling may expose errors not yet detected and deserves the analysis provided by the d-partial-area taxonomy.

If the new d-partial-area taxonomy for the Specimen hierarchy obtained as a result of the Phase 2 audit, and possibly reflecting a future release of SNOMED, were to differ meaningfully from the d-partial-area taxonomy of the 2009 release of SNOMED, then it may be advisable to reapply the auditing utilizing this new view.

2.4.3 Error Rates and the Complexity of the d-partial-area taxonomy

In Phase 1 of auditing, the bulk of the erroneous overlapping concepts and the overlapping concept errors occur for the areas {*substance*} and {*topography*}. It is interesting to compare the various ratios of errors for these two areas. The percentage of erroneous overlapping concepts in {*topography*} (61%) is about double that in {*substance*} (31%). However, when measuring the ratios of errors to overlapping concepts, the values for the two areas 0.95 and 0.89, respectively, are close. This is a result of a much higher ratio of errors to erroneous concepts for {*substance*} (2.8) than for {*topography*} (1.6). This observation indicates a correlation between the ratio of the number of errors to the number of erroneous concepts and the level of complexity of overlapping concepts, as expressed in the structure of the disjoint partial-area taxonomy. As was discussed and shown in Figure 9 and 10 in [63], the nature of the overlap is much more complex for {*substance*} with several levels in its disjoint partial-area taxonomy, while it is simpler and relatively flat for {*topography*}.

2.4.4 An Audit Report from Several Auditors

The auditing in Phase 1 was performed by two of the auditors (GE, JX) who are MDs with experience in medical terminologies. Their error report was obtained by a consensus from their individual findings. Only these errors, which were approved by Dr. Spackman, where corrected in SNOMED's July '08 release. Anecdotal evidence from the auditors was that the face-to-face consensus process seemed to follow more of a social give-and-take rather than a deep investigation about the concepts. Similar anecdotal evidence was obtained for a study of auditor performance regarding a consensus-building stage [74].

As a result, the SABOC team decided in the Phase 2 auditing to avoid the discussion-based, consensus-building effort. Instead, a combined report derived from the three auditors' Phase 2 reports is circulated. This report was anonymized and contained listings of the number of auditors for each identified error. In this second stage, each auditor was asked to indicate their agreement with each of the errors. Errors that had the support of at least one auditor were submitted to Dr. Case for further review. It seems that a second review of others' audit reports carried out by each auditor individually without the pressure of direct social interaction is functioning well in achieving an agreement level. Not only was a better level of agreement reached, but auditors backing off from certain errors were witnessed, when noticing that the other auditors did not mark them.

2.4.5 Limitations and Future Work

As can be seen from Tables 2.4 and 2.8, according to all reported measures, there is a significantly higher return for the auditing effort obtained for the overlapping concepts compared to concepts in partial-areas without overlaps. Such higher return seems to justify concentrating auditing efforts on the more complex overlapping concepts. The results confirm Hypothesis 1. More experiments with different and larger hierarchies of SNOMED and similar terminologies, e.g., NCIt [41], are needed to further confirm the finding. One idea expressed in [63] that was not confirmed by the current study was that "derived" overlapping roots (of d-partial-areas) would be more error-prone than "base" overlapping roots due to their higher complexity. The current results did not support such a phenomenon. Future studies should look again at whether this extra inherent complexity manifests itself in higher error rates in other SNOMED hierarchies.

SABOC team's interest in this research was not in studying the auditing process *per se*, but in the distribution of the unquestionable errors resulting from it. Auditor performance and the impact of various protocols in achieving better agreement among a group of auditors may be investigated in the future.

2.5 Summary

The SABOC team proceeded from the assumption that "complex" concepts warrant particular attention in quality assurance activities pertaining to SNOMED. Toward that end, an auditing methodology based on a refined abstraction network for a SNOMED hierarchy is presented, called the disjoint partial-area taxonomy, formulated in [63]. The complex concepts in this study were taken to be those residing in elements of the disjoint partial-area taxonomy that represented certain overlapping subsets of portions of a SNOMED hierarchy. These so-called overlapping concepts in the Specimen hierarchy (in two different releases of SNOMED) were identified programmatically and then put through rigorous audits. Comparing these auditing results with those from control sets, a statistically significant of higher error rate among the overlapping concepts is found. Furthermore, among the overlapping concepts, roots have a statistically significantly higher error rate than do non-roots. Thus, the auditing methodology based on disjoint partial-area taxonomy and its overlapping concepts can be seen as an important addition to the existing suite of SNOMED and SNOMED-related terminology auditing regimens.

CHAPTER 3

AUDITING SNOMED RELATIONSHIPS USING A CONVERSE ABSTRACTION NETWORK

In SNOMED, a given kind of relationship is defined between two hierarchies, a source and a target. Certain hierarchies (or subhierarchies) serve only as targets, with no outgoing relationships of their own. However, converse relationships—those pointing in a direction opposite to the defined relationships—while not explicitly represented in SNOMED's inferred view (the definition of inferred view was referred to Chapter 1 Section 1.2.2), can be utilized in forming an alternative view of a source. In particular, they can help shed light on a source hierarchy's overall relationship structure. Toward this end, an abstraction network, called the converse abstraction network (CAN), derived automatically from a given SNOMED hierarchy is presented. An auditing methodology based on the CAN is formulated. The methodology is applied to SNOMED's Device subhierarchy and the related device relationships of the Procedure hierarchy. The results indicate that the CAN is useful in finding opportunities for refining and improving SNOMED.

3.1 Introduction

A particular hierarchy may serve as a source for one relationship and the target for another. Certain hierarchies have no outgoing relationships of their own. Such a hierarchy is called a *strict target hierarchy* (or *subhierarchy*, when appropriate).

Even though a strict target hierarchy has no relationships, it does exhibit *converse relationships*—i.e., those pointing in the opposite direction to the existing incoming relationships. While these relationships are not explicitly represented in SNOMED's inferred view, available, for example, through the CLUE browser, they are, however, often utilized in data retrieval tasks or in the formation of expressions in clinical environments. They can be employed in providing an alternative view of a source hierarchy's relationship structure. A new kind of abstraction network, called the *converse abstraction network* (CAN), is introduced to represent and display a hierarchy's concepts according to their distribution of converse relationships. This network is automatically derived from the underlying inferred view of the concept network. The CAN offers a unique perspective on the source hierarchy's relationships that differs significantly from the original design view and therefore can bring unexpected structural features to light.

The SABOC team avail themselves of this unique perspective by defining an auditing methodology that utilizes the CAN and is applicable to the source hierarchy. The methodology is applied to the Device subhierarchy (of the Physical Object hierarchy) and the device relationships of the Procedure hierarchy. Potential improvements to the relationship configuration discovered through this process are presented.

3.2 Auditing Using Converse Abstraction Network

More than half of SNOMED's hierarchies are strict target hierarchies, with only incoming relationships. The Device subhierarchy is an example. The Procedure hierarchy targets it with five defining relationships: *procedure device*, *using access device*, *direct device*, *using device*, and *indirect device*. Each describes devices associated with a particular procedure. *Procedure device* subsumes the others in a role hierarchy [1]. The current analysis involves converse relationships derived from SNOMED's inferred view.

Specifically, a non-nested transform of SNOMED's original DL representation available from the relationships table is used.

3.2.1 Converse Relationship

Let us start with the definition of *converse relationship* with respect to SNOMED's inferred view. After that, the CAN is defined. Lastly, an auditing methodology is introduced based on the CAN.

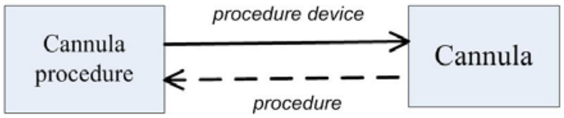


Figure 3.1 Example of converse relationship.

Consider the concepts *Cannular procedure* (from the Procedure hierarchy) and *Cannula* (from Device), connected by the relationship *procedure device* (Figure 3.1). The converse relationship of *procedure device* is defined to be the relationship that reverses its direction, connecting *Cannula* to *Cannular procedure.* In this case, it is called *associated procedure* (see the dashed arrow in Figure 3.1). A converse relationship r' will have a name derived from its original relationship r.

3.2.2 Converse Abstraction Network

As mentioned in previous chapters, various auditing methodologies are formulated based on the area taxonomy and partial-area taxonomy [2, 70]. Both are derived from a hierarchy that is the source of relationships. Please note that these taxonomies and their accompanying auditing are not appropriate for a strict target subhierarchy such as Device. In this chapter, a new abstraction network is presented that is applicable in such circumstances.

A new kind of abstraction network is defined, called the converse abstraction network (CAN), on the side of the target hierarchy of relationships in SNOMED's inferred view. Unlike the taxonomies of the previous work, the CAN is *not* a purely hierarchical structure reflecting logical concept subsumption and relationship inheritance. In fact, inheritance is not a characteristic of the converse relationships that is derived. Therefore, such inheritance is not reflected in the CAN. Moreover, the partial areas defined with respect to the CAN do not necessarily have the level of hierarchical cohesion found in the partial areas previously derived, as will be described.

Two concept groupings for the converse relationships of a SNOMED target hierarchy are defined. Let r_1', r_2',…, r_n' be converse relationships. The *area* of r_1', r_2',…, r_n' to be the set of concepts with exactly these converse relationships are defined. An area is named by its unique set of relationships (written in braces). An example is the area {*used for access by proc*, *used by proc*} ("*proc*" short for *procedure*), a set of 48 concepts from the Device subhierarchy. One of its concepts is *Endoscope*, which is a target of two relationships, *using access device* and *using device*.

It is possible that some concepts within a hierarchy are not targets of any relationships at all. For these, an additional area is defined, denoted ∅ (read "having no converse relationships"), to hold them. Collectively, the areas of a given hierarchy form a partition of that hierarchy. That is, each concept belongs to one and only one area.

The second grouping is derived from the first and is hierarchical in nature. Within an area ***A***, a concept is a *root* if none of its ancestors is also in ***A***. For each root *O* of ***A***, a set called the *partial area* containing *O* and all its descendants in ***A*** is defined. The partial area is denoted as *O*. For example, the concept *Endoscope* is a root of {*used for access by* *proc*, *used by proc*}. It and its 41 descendants (e.g., *Fetoscope*) in that area form a partial area.

In a taxonomy [2, 70], the subhierarchy residing in a partial area is completely connected. However, a partial area of a CAN may be disconnected. For example, *Ureteroscope* is in its grandparent *Endoscope*'s partial area. But its parent, *Urinary endoscope*, resides in an entirely different area, {*used for access by proc*}, thus upsetting the connectedness.

The areas and partial areas serve to give an indication of the converse relationship sources within a hierarchy and their associated hierarchical arrangement. For the purpose of visualization, a network structure based on the areas and partial areas is defined. This directed network is referred to as the *converse abstraction network* (*CAN*). Each node of the CAN represents an area. Within an area node, embedded nodes are found, each of which represents a partial area. The edges of the CAN are defined between partial areas residing in different areas as follows. Let O be a root and let P be its parent. Recall that P resides in a partial area, say, L_P that must be in an area different from O's. Then there exists an edge directed from partial area O to L_P. As examples, there are three partial areas *Urinary endoscope*, *Otoscope*, and *Rigid tracheoscope* in the area {*used for access by proc*}. The roots of the first two are children of *Endoscope*, and the root of the third is a grandchild of *Endoscope* via the parent *Rigid scope*. Thus, there is an edge from each of these three partial areas to the partial area *Endoscope*. The parent of *Endoscope* is *Scope AND/OR camera*, residing in the area ∅. As a special case, the edge in this circumstance goes from the partial area *Endoscope* to ∅. As it happens, the CAN is not a hierarchical network (e.g., a lattice).

The CAN provides a compact abstract view of the content of a hierarchy organized according to the concepts' sets of converse relationships and their IS-A arrangements. For example, there are 2,985 device concepts without any incoming relationships, and six medical balloon devices targeted by *using device* relationships.

The CAN's importance comes to light in the context of auditing relationships from one SNOMED hierarchy to another, target hierarchy. As in the previous work [2, 70], the SABOC team is looking for unexpected structural features in the CAN that could possibly be manifestations of underlying problems. For example, the concepts in ∅ have no incoming relationships whatsoever. There are also general device concepts (e.g., *Catheter*) in small partial areas having many converse relationships, while their descendants (e.g., *Vascular catheter*) appear in partial areas with fewer relationships. Such unexpected arrangements deserve attention from an auditor. In the auditing work, one needs to consider the original relationship targeting such (device) concepts and their related (procedure) source concepts. The goal in this is to find opportunities for refinement and improvement of SNOMED's relationship structure; or, in fact, to further validate the existing structure.

3.3 Results

The Device subhierarchy exhibits a total of five converse relationships, mentioned above, directed to the Procedure hierarchy. A portion of its CAN is shown in Figure 3.2. Overall, it has 22 areas and 260 partial areas. The number in parentheses in a partial area node indicates its number of member concepts. The CAN of the Device subhierarchy is not a pure hierarchical structure. In fact, one can see edges emanating from the same partial area (e.g., *Biliary T-tube*) pointing upward and downward. However, the SABOC team does lay the CAN out in levels and color-code them according to the number of relationships of the various areas. For example, the green area {*used for access by proc*} is on Level one with four partial areas (e.g., *Urinary endoscope*) and five concepts. The pink area with all five converse relationships is on Level five. It has one partial area *Catheter*. If not all partial areas are shown for an area, then the numbers of concepts and partial areas are written in parentheses. For example, {*used for access by proc*, *used by proc*} has 48 concepts and six partial areas. The largest partial area is *Endoscope* (42).

A review of Device's CAN reveals many interesting structural features, enumerated in the following. (1) The vast majority of devices (2985, 78%) are not being pointed to by any procedures. (2) An edge pointed downward may exist from a partial area with fewer relationships to a partial area with more. E.g., *Urinary endoscope* (Level 1) has an edge to *Endoscope* (Level 2). (3) Many partial areas are singletons, meaning they contain only one concept each. (4) Some small partial areas are of a very broad nature, such as *Catheter* and *Drain*. (5) Certain partial areas include extremely high-level, non-specific devices, such as *Device* itself, which subsumes all the devices in all the CAN's partial areas. (6) Certain partial areas are pointed to by one or very few procedures. (7) Devices of a similar nature, such as *Venous catheter* and *Arterial catheter*, reside in different areas.

These features were used to focus the auditing efforts on certain concepts and relationships of the Procedure hierarchy (targeting Device), and thus provided opportunities to find potential errors that would probably not be detected directly from the Procedure hierarchy. In the following discussion, examples are provided pertaining to these observed features and review their value as indicators of potential errors or improvements in the modeling.

3.4 Discussion

The CAN exhibits properties that differentiate it from the previous partial area taxonomies [2, 70]. For example, with partial area taxonomies, there is inheritance of relationships among partial areas along the *child-of* hierarchy. No such inheritance is guaranteed for the CAN. For example, *Vascular Catheter* is in the red area {*used for access by proc*, *acted on directly by proc*, *used by proc*} and has an edge directed to the *Catheter* partial area (in the pink, Level-five area). Two of the relationships are not appearing for *Vascular Catheter*. In the figure, this is manifested by the edge pointing downward, while in the partial-area taxonomy the *child-of* relationships point up to areas with fewer relationships.

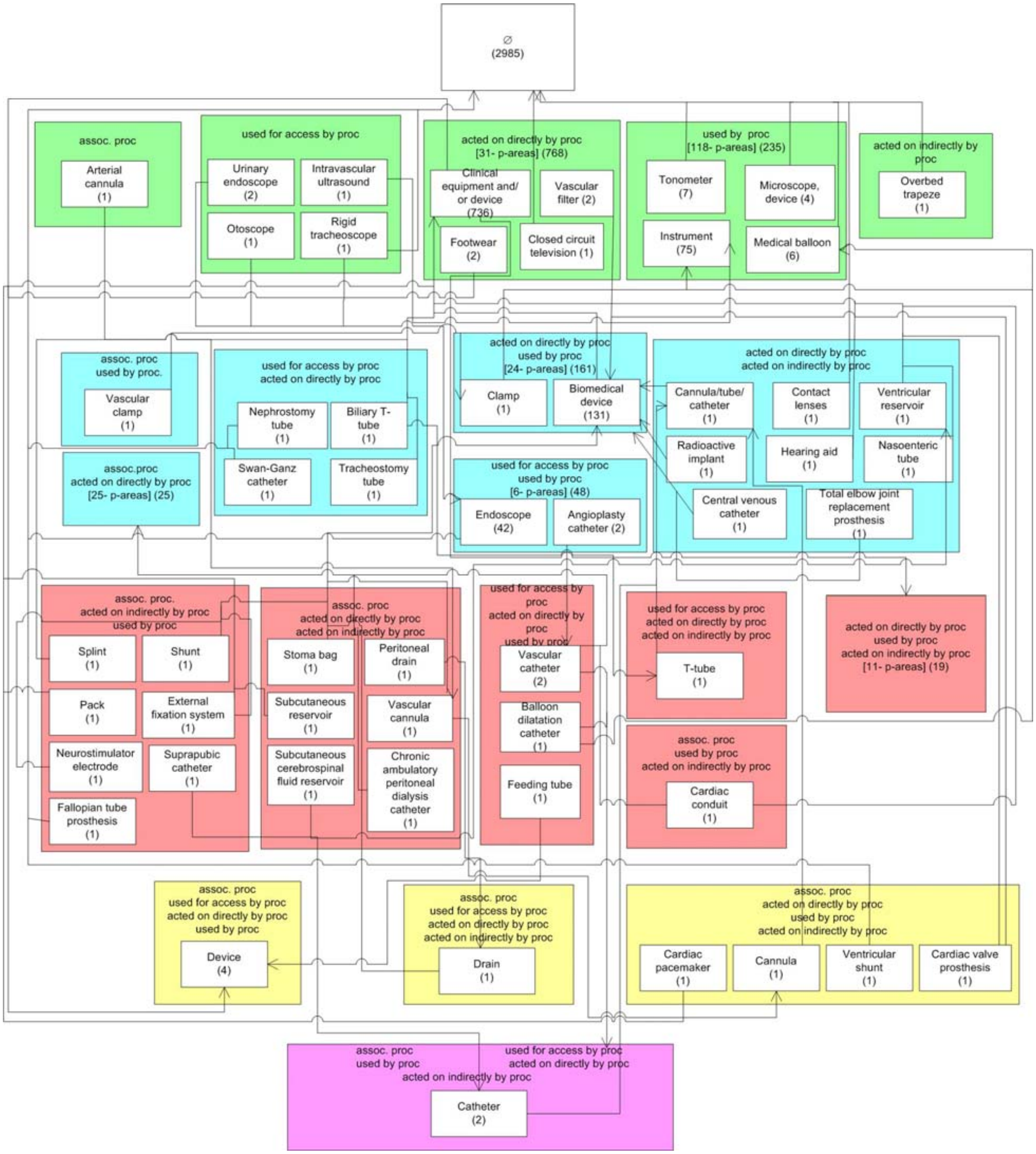


Figure 3.2 CAN of the Device sub-hierarchy.

Intracavitary brachytherapy is a procedure that does not have any device relationship. However, the procedure achieves sufficient definition by using another relationship, *method* to *Brachytherapy – action*. Due to this, *Brachytherapy implant* does not have any incoming relationships and resides in ∅.

Additionally, the Device subhierarchy does not use inheritance of attribute values but instead relies on the notion of *refinability*. A refinability value is assigned to every relationship type between a pair of concepts, usually at some ancestral level. As a result, many descendant concepts will reside in ∅. For example, *Charnley total hip prosthesis* resides in ∅ along with its siblings. However, the procedure *Total hip replacement* points to the parent device, *Total hip replacement prosthesis*. Thus, the procedure achieves sufficient definition while allowing the device to be refined as needed by the procedure's descendants.

However, from a user perspective, as in a decision-support system or other terminology-driven systems, such an arrangement may be perceived as deficient. If one wants to select a specific endoscope for a gastrointestinal procedure while that procedure is sufficiently defined with the device *Endoscope*, one may be able to select, say, *Otoscope* as the device. And since the Device subhierarchy does not have any outgoing relationships, the devices cannot be defined by the body systems or organs they act upon.

The CAN also highlights the fact that a partial area may have a downward edge directed to another partial area with fewer relationships. For example, *Urinary endoscope*, its child *Nephroscope*, and its sibling *Otoscope* reside on Level 1. However, all three are children/descendants of *Endoscope* and have other siblings that reside on Level 2 along with *Endoscope*. Moreover, *Urinary endoscope* and *Otoscope* are each pointed at by only one procedure. With taxonomies, such small partial areas are seen as being indicative of possible errors, and one might expect that a child have at least as many device relationship types as its parent. Regarding *Urinary endoscope*, it might be more appropriate for urinary procedures with relationships currently using *Endoscope* to have the more specific target, instead. This would result in *Urinary endoscope*'s movement into the *Endoscope* partial area. However the current structure is still sufficient by SNOMED criteria.

While the notion of sufficient definition may explain the use of higher-level device categories, some may seem at too high a level. For example, *Removal of Kantrowitz heart pump* points to *Device* via *direct device*. While acknowledging refinability, this assignment seems overly general since *Device* roots a significant subhierarchy. As is the case with other fully specified procedure concepts, such as *Open insertion of Hickman central venous catheter*, the procedure should point to either *Heart pump* or the more specific *Kantrowitz heart pump*. However, these device concepts do not exist in SNOMED. Their omission suggests a needed refinement.

The Hickman example highlights another observation. While the fully specified Hickman procedure above uses the explicitly specified *Hickman catheter* device, its sibling, *Open insertion of Broviac central venous catheter*, does not. The Broviac device is missing. In this case, two "parallel" concepts are modeled differently and offer an opportunity for further refinement.

In this discussion, an effort is made to illustrate various kinds of problems exposed by the alternative view offered by the CAN. Unlike SABOC team's previous work, this study did not unearth a large number of errors. This is not surprising since this part of SNOMED received comprehensive scrutiny by its editors. This is one possible (and potentially the most preferred) result of an auditing effort. However, it is better to emphasize that the CAN is not proposed as an all-inclusive auditing method but rather as an additional tool in an auditor's toolbox. The abstraction view is structural-based and helps expose anomalies that might not be uncovered otherwise. It is complimentary to other methods such as DL-based auditing methodologies.

3.5 Summary

Converse relationships, derived from relationships in SNOMED's inferred view, have been used in the construction of a new kind of abstraction network, the CAN, for a strict target hierarchy. An auditing methodology for such a hierarchy's incoming relationships whose basis is the CAN was presented. The results of applying this methodology to the Device subhierarchy indicate that the CAN is a useful auditing vehicle that can bring various aspects of the relationship structure to light and aid an auditor in refining and improving SNOMED.

CHAPTER 4

USING AN ABSTRACTION NETWORK IN COMPLEMENT TO DESCRIPTION LOGICS FOR QUALITY ASSURANCE IN SNOMED

The objective of this research is to investigate errors identified in SNOMED CT by human reviewers with help from the abstraction network methodology and examine why they had escaped detection by the DL classifier. Two examples of errors are presented in detail (one is missing IS-As and the other is duplicate concepts). After correction, SNOMED is reclassified to ensure that no new inconsistency was introduced. Towards the end, a conclusion is drawn that DL-based auditing techniques built in terminology development environments ensure the logical consistency of the terminology. However, complementary approaches are needed for identifying and addressing other types of errors.

4.1 Introduction

Modern terminologies including SNOMED and the NCI Thesaurus are created with the support of DLs, which ensures the logical consistency of the terminological assertions. However, errors still exist in SNOMED, even after the automatic classification of the DL reasoners. As presented in previous chapters, the types of errors include inaccurate or incomplete logical definitions (e.g., errors in the nature or in the target of asserted relationships, as well as missing relationships). Chapter 1 Section 1.2.5 discussed a number of auditing techniques based on lexical, structural, and ontological principles. In general, those approaches applied computational method to the identification of potential errors. The automated processes are designed to facilitate the work of human editors (subject matter experts) and it contributes to the quality assurance of biomedical terminologies.

The objective of this research is to investigate errors identified in SNOMED by human reviewers, with help from the abstraction network methodology. More specifically, the reason why such errors could not be identified by a DL classifier is examined and a strategy for using the abstraction networks in complement to DL-based techniques is proposed for the quality assurance purposes. The contribution of this work is not to propose novel approaches to identifying errors in SNOMED, but rather to tease out differences between existing approaches based on several cases of errors thoroughly investigated.

4.2 Comparison of Description Logics with Abstraction Network for Quality Assurance

4.2.1 Description Logic for Quality Assurance

Description logics (DL) are a family of knowledge representation formalisms often used as ontology languages [75]. Not only does DL provide support for defining concepts, but it also provides methods for reasoning about concepts and their instances. DL reasoning services are carried out by DL classifiers.

The basic inference on a concept expression is subsumption, i.e., comparing two classes and checking whether one class is more general than the other. For example, *Brain disorder* is more specific than (i.e., is subsumed by) *Disorder*, because *Brain disorder* is defined as a disorder located to the brain. Another important inference is concept satisfiability. A class is deemed unsatisfiable (i.e., inconsistent) if it cannot possibly have any instances. For example, nothing can be at the same time a procedure and an anatomical structure. If a class **C** were defined as a subclass of both Procedure and Body structure, while Procedure and Body structure are defined to be disjoint, a DL classifier would identify **C** as unsatisfiable. The interested reader is referred to [75] for additional details about DL.

There are, however, many different dialects of DL in terms of the set of constructors they offer, resulting in different levels of expressiveness for what can be defined. The expressiveness of the DL also determines the kinds of inference a DL classifier is enabled to perform and the kinds of logical inconsistency it is able to identify. The dialect of DL natively used by SNOMED is "*EL*", whose expressiveness is relatively limited. For example, *EL* does not allow disjunction to be stated between classes and the example of unsatisfiability presented earlier could therefore not be identified by the DL classifier used for the creation of SNOMED.

From the perspective of error identification in ontologies, two major types of errors can be distinguished. Type I errors are the logical inconsistencies in concept expressions that can be detected by DL classifiers (assuming the DL dialect used is expressive enough to state the circumstances under which concepts would be inconsistent, e.g., disjointedness). In contrast, Type II errors are those content errors (e.g., incorrect relationships, missing relationships) that would not generate logical conflicts in the DL system. Quality assurance processes in SNOMED ensure that all Type I errors have been identified and corrected before the terminology is released to users. All the errors under investigation in this study are therefore Type II errors. (Here, Type I and

Type II errors are defined in reference to the level of expressiveness of the *EL* dialect of DL).

As mentioned earlier in Chapter 1 (Section 1.2.2), the inferred view is automatically derived from the asserted view by a DL classifier. In this research, the inferred view is analyzed, but, unlike most users, we also modify the asserted view and use a DL classifier in order to check any suggested changes for consistency.

4.2.2 Abstraction Network for Quality Assurance

In Chapter 1, two abstraction networks: *area taxonomy* and *partial-area taxonomy* have already been defined. Both are derived automatically. These two abstraction networks are structural methodologies developed for reducing the complexity of large biomedical terminologies [2]. The structural methodologies that applying the two abstraction networks are called Abstraction Network (AN) methodologies. As mentioned earlier, the AN methodology is based on the relationships and their inheritance patterns in the hierarchies of the terminology. It has been applied to auditing SNOMED. Here, a brief description of its underlying principles is given and its application to SNOMED is reviewed. The examples in the research focus on the Specimen hierarchy of SNOMED.

AN provides an abstraction of the hierarchical and relationships of concepts in a SNOMED CT hierarchy. The idea is to partition such concepts into structural uniformity groups (strUGs), and then to refine the partition into semantic uniformity groups (smtUGs). A detailed description can be found in [2, 68, 70]. In previous chapters, *structural uniformity group* is referred to as *area*, while *semantic uniformity group* is referred to as *partial-area*. For the sake of clear demonstration, strUG and smtUG are used in this chapter.

A ***structural uniformity group (strUG)*** is the group of all concepts with exactly the same set of relationships. In a graph structure, a node to represent a strUG is used. The label for the strUG node is the set of relationships in which its concepts participate.

Five different relationships are introduced to the concepts of the Specimen hierarchy; they are *substance*, *morphology*, *procedure*, *topography*, and *identity*[1]. For example, the concept *Surgical excision sample* has one relationship *procedure* pointing to a concept *Excision* (from the Procedure hierarchy). Therefore, the concept *Surgical excision sample* is in the strUG{*procedure*}. Similarly, the concept *Abscess swab* has two relationships *procedure* and *morphology* pointing to *Taking of swab* and *Abscess morphology* (from the Procedure and Body structure hierarchy, respectively). Thus, *Abscess swab* is in the strUG{*procedure*, *morphology*}. Note that strUGs do not overlap, because, by construction, one given concept belongs to one and only one strUG corresponding to its relationship pattern. Therefore, the entire set of strUGs forms a partition of the concepts in a given hierarchy of SNOMED.

StrUGs can be organized into a graph structure. Hierarchical relations between strUGs are determined by the inclusion of the sets of relationships they represent. For example, the strUG{*procedure*} subsumes the strUG{*procedure*, *morphology*}. Figure 4.1(a) shows a portion of the graph of strUGs for the Specimen hierarchy. Each colored box represents a strUG. The boxes are color-coded to differentiate the levels. Each level

---

[1] The full names of these relationships are *specimen substance*, *specimen source morphology, specimen procedure, specimen source topography*, and *specimen source identity*, respectively.

corresponds to the number of relationships in the strUG. The concepts in the strUG Ø have no relationships.

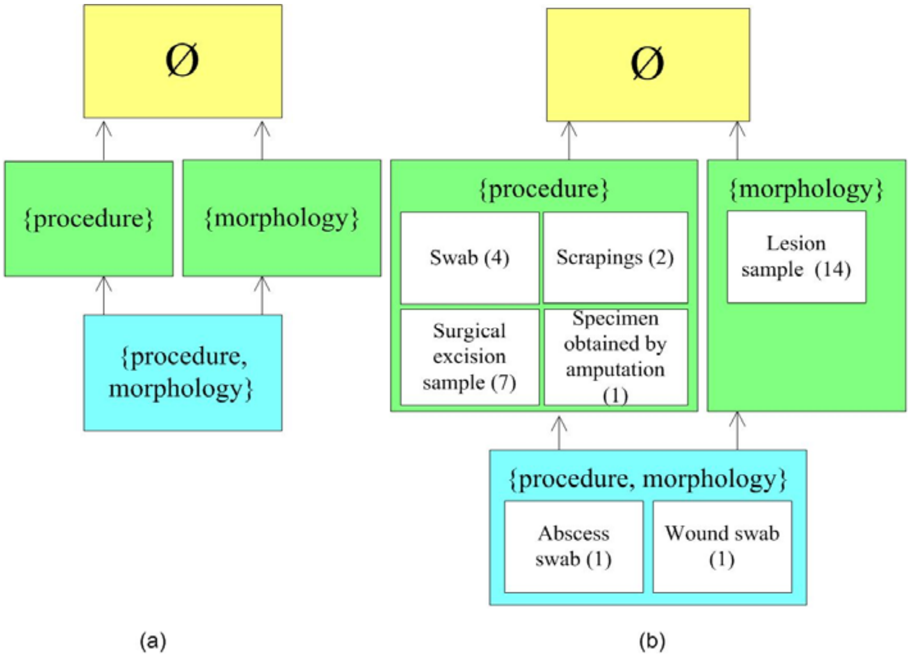


Figure 4.1 (a) Portion of the graph of StrUGs for the Specimen hierarchy (b) Corresponding portion of the graph of smtUGs.

A ***semantic uniformity group (smtUG)*** is a group of concepts within a structural uniformity group sharing the same lowest common ancestor (LCA). In other words, the smtUG groups concepts with the same relationships by hierarchical relationships. The label for the smtUG is the LCA from which all other concepts in the smtUG are descendants. A strUG may have more than one LCA and thus more than one smtUG. The smtUGs form a semantic subdivision of the strUG, but not necessarily a subpartition of it, since a concept may have more than one LCA.

The graph of strUGs in Figure 4.1(a) can be refined with the smtUGs contained within each strUG, as shown in Figure 4.1(b). For example, the strUG{*procedure*} contains the four smtUGs: smtUG(*Swab*), smtUG(*Scrapings*), smtUG(*Surgical excision* *sample*) and smtUG(*Specimen obtained by amputation*). The number in the parentheses indicates the number of concepts within a smtUG. For example, in the smtUG(*Surgical excision sample*), there is a total of seven concepts. The six hidden concepts are all subsumed by *Surgical excision sample*.

The strUGs and smtUGs hide some of the complexity of the terminology. In previous chapters, this abstracted view has already proved itself to be a useful auditing tool for manual review of biomedical terminologies by subject matter experts.

4.2.3 Review of Auditing Methods Using Abstraction Networks

Several strategies have been devised to help subject matter experts review parts of SNOMED based on the Abstraction Network methodology.

Group-based auditing takes advantage of the grouping of concepts in semantic uniformity groups [2]. All concepts from a given group are reviewed at the same time, making it easier for experts to identify discrepancies among concepts expected to be both structurally and semantically similar. Errors exposed via group-based auditing include redundant concepts, erroneous relationships, incorrect *IS-A* assignments, and other content errors.

Auditing "complex" concepts focuses on those concepts within a structural uniformity group, which belong to several semantic uniformity groups because they have ancestors in several smtUGs [68]. Errors found in such complex concepts include missing children and incorrect parents.

Error concentration based auditing is predicated on the fact that small semantic uniformity groups are more likely to contain errors, because small sets of similar concepts might have received less modeling attention, compared to larger sets (e.g., based on a concept model). The correlation between small smtUG size and error concentration was assessed in [70].

4.3 Case Study

Two of the errors detected in SNOMED by subject matter experts with help from the Abstraction Network methodology were selected and reported to the International Health Terminology Standards Development (IHTSDO)[2], the organization in charge of SNOMED. The objective in this chapter is to investigate these cases and examine how they escaped detection by the DL classifier used to check the logical consistency of SNOMED.

DL reasoners are stand-alone tools that point out logical inconsistencies in an ontology. In contrast, the Abstraction Network methodology helps organize the workflow of subject matter experts, in order to focus their attention to parts of the ontology where errors are likely and by grouping the concepts to be audited according to the principles described earlier.

The two errors under investigation were identified in the Specimen hierarchy of SNOMED. In the first one, “amputation,” it was argued that two sibling concepts actually stand in a subsumption relationship. The issue is thus a missing IS-A between these two concepts. The second case, “leukocyte,” highlights two concepts that are arguably equivalent but stand in an IS-A.

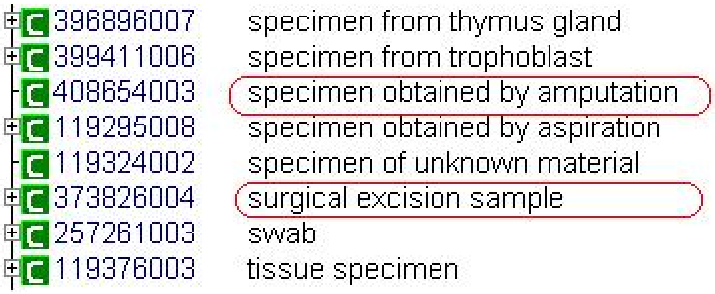
---

[2] http://www.ihtsdo.org/

In addition to discussing the errors, it is also desirable to test the remediation suggested to the IHTSDO. Toward this end, the asserted version of SNOMED is loaded in OWL DL into the ontology editor Protégé[3] and tested the suggested changes with the DL classifier Fact++[4]. The goal is to verify that the proposed changes did not introduce any inconsistencies into SNOMED. Classification was performed on a standard desktop machine with the 64-bit Microsoft Windows operating system and 4 GB of RAM. The classification of the OWL version of the SNOMED CT takes about 17 minutes.

4.3.1 Case 1: Amputation

This error was identified by the subject matter expert while examining a group of concepts from the Specimen hierarchy corresponding to one particular structural uniformity group, namely, strUG{*procedure*}. By construction, the concepts naming the smtUGs within a strUG are not expected to stand in any kind of hierarchical relationship. The assumption for the subject matter expert reviewing the concepts from a strUG is that they are all expected to be siblings. Therefore, reviewing these concepts as a group makes it easy to identify errors including missing or incorrect parent/child relationships, for example.



396896007	specimen from thymus gland
399411006	specimen from trophoblast
408654003	specimen obtained by amputation
119295008	specimen obtained by aspiration
119324002	specimen of unknown material
373826004	surgical excision sample
257261003	swab
119376003	tissue specimen

Figure 4.2 "Specimen obtained by amputation" and "Surgical excision sample" displayed in the CliniClue browser.

---

[3] http://protege.stanford.edu/
[4] http://owl.man.ac.uk/factplusplus/

Figure 4.2 shows a portion of the inferred view of SNOMED displayed in the CliniClue browser[5]. The two concepts circled in red, *Specimen obtained by amputation* and *Surgical excision sample*, are siblings. Both of them are in the Specimen hierarchy under the root concept *Specimen*. The corresponding target concepts with the relationship procedure are *Amputation* and *Excision*, respectively, in the Procedure hierarchy, under the parent concept *Surgical removal* (not shown in the figure). The four concepts *Specimen obtained by amputation*, *Surgical excision sample*, *Amputation*, and *Excision* are fully defined.

The subject matter expert determined that *Specimen obtained by amputation* is, in fact, a kind of *Surgical excision sample*. The fact that the two concepts were grouped in strUG{*procedure*} made it easier for the expert to identify this error. Of note, there was no logical inconsistency in the concept expression, and the DL reasoner failed to detect the missing subsumption relationship because its absence did not create any kind of conflict in the terminology. One particular reason why no conflict could be identified is because there was a parallel error on the target side. The target concepts *Amputation* and *Excision* are siblings (descendants of *Surgical removal*), while amputation is actually a kind of excision. Because of a missing IS-A in parallel on both sides of the procedure relationship, there was no logical error that could be identified by the DL classifier.

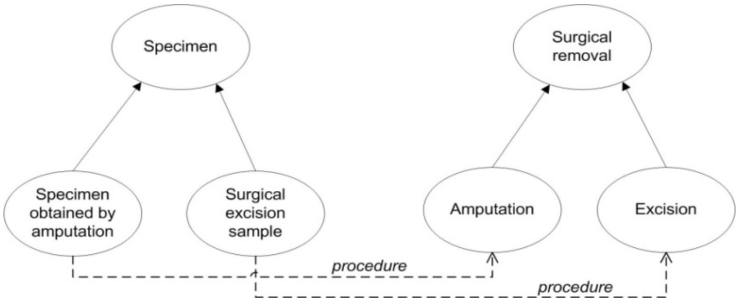
From the perspective of the Abstraction Network, both smtUG(*Surgical excision sample*) and smtUG(*Specimen obtained by amputation*) are in the strUG{*procedure*} (see Figure 4.1(b)). But the existence – indicated by the expert – of an IS-A between these two

---

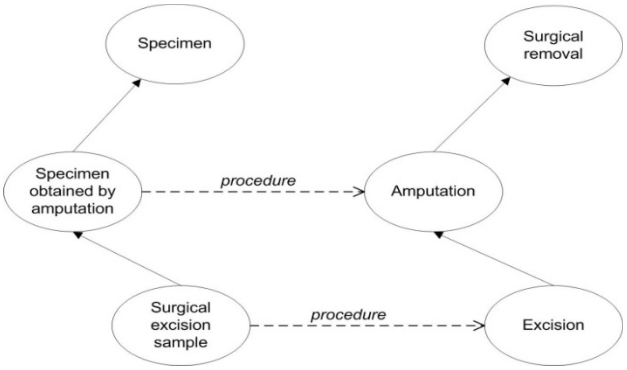
[5] http://www.cliniclue.com/

concepts within the same strUG{*procedure*} violates the principles under which the strUG was constructed.

Figure 4.3 shows the comparison before and after addition of the missing IS-As. As a result of this modification, *Specimen obtained by amputation* is now subsumed by *Surgical excision sample*, and smtUG(*Surgical excision sample*) has gained a new member.



(a)



(b)

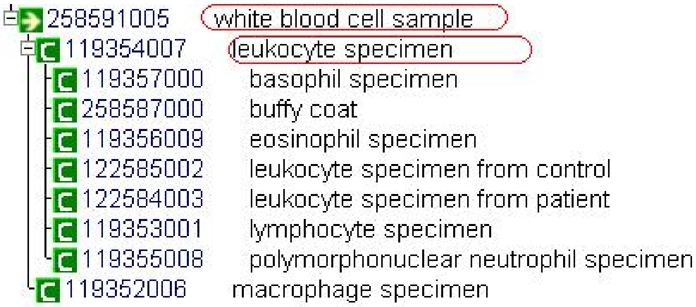
Figure 4.3 Parent-child error with *Surgical excision sample* and *Specimen obtained by amputation* (a) Before correction (b) After correction.

The target hierarchy (Procedure) is modified by making *Surgical Excision* the superclass of *Amputation* in the authors' copy of SNOMED in Protégé, while leaving the source hierarchy (Specimen) unchanged. After reclassification, the classifier is seen to have used the changes made to the target hierarchy (Procedure) to automatically make parallel changes to the source hierarchy (Specimen), where *Surgical excision sample* has become the superclass of *Specimen obtained by amputation* (Figure 4.3(b)).

4.3.2 Case 2: Leukocyte

This error was identified by the subject matter expert while examining a group of concepts from the Specimen hierarchy corresponding to one particular semantic uniformity group, namely, smtUG(*White blood cell sample*). By construction, concepts within a smtUG are expected to stand in an IS-A with the lowest common ancestor after which the smtUG is named. The assumption for the subject matter expert reviewing the concepts from a strUG is that they are all expected to be distinct and descendants of *White blood cell sample*. Therefore, reviewing these concepts as a group makes it easy to identify duplicate concepts, for example.

As shown in Figure 4.4, *Leukocyte specimen* is one of the children of *White blood cell sample*. The subject matter expert determined that *Leukocyte specimen* and *White blood cell sample* are, in fact, duplicate concepts. The fact that the two concepts were grouped in smtUG(*White blood cell sample*) made it easier for the expert to identify this error.



- 258591005 white blood cell sample
  - 119354007 leukocyte specimen
    - 119357000 basophil specimen
    - 258587000 buffy coat
    - 119356009 eosinophil specimen
    - 122585002 leukocyte specimen from control
    - 122584003 leukocyte specimen from patient
    - 119353001 lymphocyte specimen
    - 119355008 polymorphonuclear neutrophil specimen
  - 119352006 macrophage specimen

Figure 4.4 *Leukocyte specimen* and *White blood cell sample* displayed in the CliniClue browser.

In DL, concepts exhibiting the same logical definitions are treated as equivalent concepts by the classifier. In this case, the DL classifier did not identify these two concepts as equivalent, because the logical definitions were actually slightly different. *Leukocyte specimen* is a primitive concept, whereas *White blood cell sample* is fully defined. Because the definition of *Leukocyte specimen* is underspecified (primitive), the DL classifier cannot recognize it as equivalent to the fully defined *White blood cell sample*.

From the perspective of the Abstraction Network, there is no difference between primitive and defined concepts. Only the set of relationships is taken into account during the creation of the groups.

The definition of *Leukocyte specimen* is modified in the authors' copy of SNOMED in Protégé, so as to make it fully defined instead of primitive. After reclassification, *White blood cell sample* and *Leukocyte specimen* were indicated as being equivalent concepts.

4.4 Discussion

4.4.1 Strengths and Limitations of Each Approach

The main advantage of DL is that it identifies errors completely automatically, while the Abstraction Network (AN) methodology only constrains the workflow of subject matter experts in such a way that it facilitates their work and improves their chances of identifying errors by reducing the complexity of the terminology and by organizing the concepts to be reviewed in small groups, with assumed relationships among concepts within and across groups.

Unlike the DL classifier, the AN does not rely on defined concepts, but simply takes advantage of the structural properties of concepts, i.e., their sets of relationships. Unlike AN, the DL classifier processes the terminology as a whole and can address remote inconsistencies, whereas experts tend to focus on a small portion of the terminology and may not foresee the consequences of local changes to distant parts of the terminology.

Finally, DL classifiers are limited to the identification of logical inconsistencies. Moreover, they are limited in the type of logical inconsistencies they can identify by the level of expressiveness of the dialect of DL used for creating the ontology [76]. In contrast, subject matter experts guided by the Abstraction Network methodology can address a wider range of issues (i.e., beyond logical inconsistencies) and identify content errors, such as inaccurate and missing relationships.

4.4.2 Auditing Strategy

The DL classifier is used for detecting logical inconsistencies at the time the terminology is built. The performance of the classifiers has improved tremendously in the past few years and the editors of large terminologies will soon enjoy real-time classification. The use of the Abstraction Network methodology is recommended for targeted auditing, as a possible alternative to dual editing. However, multiple auditing strategies combining lexical, structural, and ontological methods are required for quality assurance of large, complex terminologies such as SNOMED.

4.4.3 Current Developments and Future Work

One of the limitations of the Abstraction Network methodology is that it relies heavily on the structure of relationships of the concepts and is therefore not applicable to concepts with few or no relationships. In order to address this limitation, the converse abstraction network [77] is developed as described in Chapter 3.

4.5 Summary

In this work, the differences between two approaches are examined to identifying errors in large biomedical terminologies such as SNOMED. On the one hand, DL classifiers can automatically identify logical inconsistencies in the terminology. On the other, the Abstraction Network methodology helps experts perform targeted manual reviews of the terminology by reducing its complexity and grouping the concepts by their structural and semantic properties. The differences between the two approaches are illustrated through two cases of errors identified in SNOMED.

CHAPTER 5

AUDITING AND COMPLEXITY OF SNOMED TAXONOMIES

SNOMED CT is an extensive terminology with an attendant amount of complexity. Two measures are proposed for quantifying that complexity. Both are based on abstraction networks, called the area taxonomy and the partial-area taxonomy, that provide, for example, distributions of the relationships within a SNOMED hierarchy. The complexity measures are employed specifically to track the complexity of the versions of the Specimen hierarchy of SNOMED before and after it is put through a sequence of auditing processes. The complexity measures for the pre-audit and post-audit versions are compared for two periods of auditing efforts for the Specimen hierarchy. The first audit's results have been previously published elsewhere. The second audit comprises two separate efforts. In the first, designed especially for this study, auditing of a special class of "singleton concepts" (defined with respect to the partial area taxonomy) is conducted. In the second, conducted during the same period, another class of concepts called "overlapping concepts were audited (with results reported elsewhere). The complexity results show that the initial auditing processes lead to a simplification of the terminology's structure. The results for a further auditing process are mixed.

5.1 Introduction

In [2, 70], the SNOMED's lateral relationships (also called "attribute relationships") and their inheritance patterns within hierarchies were utilized to formulate structural methodologies for auditing SNOMED. These methodologies utilize two abstraction networks, the *area taxonomy* and the *partial-area taxonomy*, that serve to capture the structure of a hierarchy in a compact manner. The taxonomies highlight where errors tend to concentrate [70] and offer techniques to detect them [2]. The errors reported in [2, 70] were subsequently corrected in later releases of SNOMED.

In this research, the impact of the corrections of errors on the complexity of a SNOMED hierarchy is investigated. The hypothesis is that, in general, errors contribute to structural disorderliness. The question is: can one expect to see a simplification of the hierarchy structure due to the reduction of such disorderliness after an auditing regimen has been carried out? For this, one needs to posit a way to assess the complexity of a hierarchy. The previously defined taxonomies offer a natural framework for this, since they are derived via structural analyses of the underlying hierarchy. The proposed assessment measures are applied to the Specimen hierarchy in order to track changes in its complexity over two periods of auditing efforts. This is done for the July 2004 version (the one prior to the corrections), the July 2007 version (after the corrections), and the July 2008 version (after further auditing in 2007). One might expect that the 2007 Specimen hierarchy is error free since it underwent several comprehensive audits in 2004. But that is too optimistic. Some errors that were hidden before were then exposed by changes made during the process of correction. Also, a few new concepts had been added to the hierarchy in the interim, and their introduction may have indeed led to new errors. Furthermore, both editing and auditing of a hierarchy are difficult tasks which by themselves are never foolproof. An auditor may very well overlook some errors, and the editorial policies may be incomplete or inconsistent.

Thus, in the process of doing the complexity assessment, an audit is performed for all the partial-area taxonomy's foundational concept groups (called *partial-areas*) of the consisting of just one concept (so-called "singletons") with respect to the July 2007 release, based on the methodologies [2, 70]. The errors found are reported in this research. Further results of auditing another class of concepts (called *overlapping* concepts) [59] are also applied. These extra audits provide us with a third version of the hierarchy on which to assess the complexity measures and a second audit period from July 2007 to July 2008 to track changes in complexity due to the second auditing effort. Any further impact of this subsequent audit effort on the complexity measures will be looked for, in comparison to the impact of the initial audit effort for the same hierarchy. An initial report of this study appeared in [62]; however, the research further evolved including changes in the definitions of the complexity measures.

5.2 Methods

The issue investigated in this research is how to measure the complexity of a SNOMED hierarchy. In particular, it is interesting to study how complexity measures reflect on the results of auditing tasks performed on a given hierarchy following its evolution over multiple releases. One natural criterion is a global weighting function for a hierarchy such as size (the number of concepts) or height (number of levels in the longest hierarchical path). Indeed, in a comparison of such measures following the first audit of the Specimen hierarchy in the 2004 SNOMED release, the number of concepts was reduced from 1,056 to 1,044 (July 2005 release), and the height was reduced from 12 to ten. At the same time, SNOMED's total concepts went up from 357,134 to 364,461. Furthermore, only two hierarchies of SNOMED decreased in size during this period, the second of which was the huge Clinical Finding hierarchy. The author attributes the decrease in the size of the Specimen hierarchy, which went against the general trend of growth in SNOMED during the same period, to the correction of duplicate concept errors (such as *Ear sample* and *Specimen from ear* [2]) and the removal of improper concepts due to the auditing efforts[2, 70]. The former were caused by the failure to identify the synonymy of "sample" and "specimen" when integrating SNOMED RT and CTV3 into SNOMED CT [78]. The errors found were reported to College of American Pathologists (CAP) by K. Spackman and were implemented in future releases. The reduction in height can be attributed to finding errors in some of the most complex concepts in the hierarchy, which participated in the longest hierarchical paths.

However, these measures are more magnitude measures than complexity measures. The size measure accounts only for limited auditing impacts such as erroneous concepts eliminated from the hierarchy, but not for other errors that were corrected. The size is also influenced by concepts added to the hierarchy as part of normal modeling expansion. The height measure reflects only auditing of a few concepts in the longest hierarchical path. Furthermore, such global measures fail to take into account the role of lateral relationships in the complexity of the concepts. For example, a hierarchy may keep its size and height while going through an auditing process, which may make it simpler or more complex.

To illustrate the difficulty of using the size measure, note that the Specimen hierarchy grew from 1,044 concepts in July 2005 to 1,052 in January 2007, while no special auditing was applied. Finally, the number grew to 1,056 (the original number in

2004) in July of 2007, before a second auditing effort was applied. Having two releases (2004 and 2007) of the same size does not necessarily imply they are of the same complexity.

As a more appropriate way of measuring the complexity of a hierarchy, it is suggesting to utilize the area taxonomy and partial-area taxonomy. The area taxonomy reflects the lateral relationship (just "relationship," for short) distribution of all the concepts in the underlying SNOMED hierarchy. The partial-area taxonomy further shows hierarchical cohesiveness [52], where concepts subsumed under a common root concept are clustered into a partial-area. All these concepts elaborate the semantics of their root. Thus, the partial-area taxonomy can support a measure of hierarchical complexity.

5.2.1 Structural Complexity Measures

The author asserts that a concept C with two given relationships is more complex than a parent concept P exhibiting only one of those relationships, since concept C with multiple relationships expresses more detailed knowledge than concept P. Similarly, the author asserts that a concept C with three given relationships is more complex than a parent concept having two out of the three relationships.

For example, as is seen from Figure 1(a) and Figure 1(c), the concept *Skin swab* in the area {*procedure*, *morphology*} has a parent *Swab* in the {*procedure*} area, and another parent *Dermatological sample* in the {*morphology*} area. In this case, *Skin swab* is the specialization of the two parents. From the complexity point of view, it is more complex as compared to either one of its parents because it has the extra knowledge expressed by the relationship inherited from the other parent.

Similarly, a root concept *Skin ulcer swab* in the area {*morphology*, *topography*, *procedure*} (see Figure 1(a) and Figure 1(c)) has three parents in three separate areas. One parent *Skin swab*, with two relationships, was mentioned above. Another parent *Skin lesion sample*, with two relationships, is from {*morphology*, *topography*}. A third parent *Specimen from ulcer*, with the one relationship *morphology*, is a non-root concept in the partial-area *Lesion sample* in {*morphology*}.The concept *Skin ulcer swab* is more complex than the one-relationship parent *Specimen from ulcer*. It is also more complex than each of the two parents with two relationships, because of its extra relationship.

The higher complexity when comparing a descendant concept to its ancestor is obvious. In general, the structural complexity of a concept by the number of its relationships is measured, since as mentioned, the structure of a concept is its set of relationships. In the context of the area taxonomy: a concept on a lower-numbered level is simpler than a concept on a higher-numbered level. This assumption is called the *structural assumption*, since it is based on a structural feature of the area taxonomy. In measuring structural complexity by the number of relationships, independent of their kind, the author extends the notion of higher structural complexity, from the case of comparing a child concept to its parent concept, to the case of comparing any pair of concepts, where the first has more relationships than the second. The justification for this generalization is that, even in the first case, the reason for the higher complexity of the child is its extra relationship. The area levels of the area taxonomy serve to partition the concepts of the hierarchy according to their numbers of relationships, and thus partitioning the concepts according to their structural complexity. If, as a result of an auditing phase, one sees an increase in the number of concepts in a lower-numbered area level of a hierarchy at the expense of a decrease in the number of concepts in a higher-numbered area level of this hierarchy, then this change can be interpreted as a simplification of the hierarchy structure. Such a change may occur when discovering an unnecessary relationship for a group of concepts. Of course, a concept must first be modeled with all its necessary relationships. A simpler representation of a concept is seen as a desired quality in the modeling of a terminology, but is only secondary to correctness. Hence, the auditing process should not seek to delete required relationships just for the sake of simplification. However, as a result of auditing, where relationships of concepts are removed or added, it is expected to see changes in the structural complexity.

Hence, one is looking for a complexity measure that will enable the comparison of two states of the same SNOMED hierarchy as it evolves over time. It is interesting in a measure that reflects the number of concepts in the various levels and the changes to those numbers due to the migration of concepts from one level to another, (as their relationships change) as a result, for example, of auditing. This is different from, say, a global complexity measure that just reflects the total number of relationships in a hierarchy and their partition into levels. Such a related measure is of course also important and will be introduced later.

To formalize this measure, the structural complexity function S(x, H) is defined, which is a function from the non-negative integers of a hierarchy to the number of concepts with the corresponding number of relationships, where x represents a level and H is a hierarchy. That is, S(x, H) is the number of concepts on Level x of the area taxonomy of hierarchy H. When there is no ambiguity regarding H, it will often be omitted. For example, according to the previously defined area taxonomy in Figure 3, Level 0 has 21 concepts, so S(0) = 21. For a given hierarchy of Levels 0, 1, 2, …, *m*, the sequence is often written as (S(0), S(1),…, S(m)).The function S is a structural measure as it depends solely on the number of relationships, not on their kind. It is a global structural measure for the complexity of the hierarchy because it is dependent on all concepts and their respective structure. To interpret the structural complexity: if more concepts lose relationships than gain relationships, then in the area taxonomy, there is an increase in the number of concepts in a lower-numbered levels and a decrease in the number of concepts in higher-numbered levels. In this case, one can say that the complexity function is going through a downward weight-shifting towards the lower-numbered levels; one can say that the structural complexity is reduced.

The Aggregated Structural Complexity S_A is also defined on a hierarchy that has the compliment to one of the reciprocal of the average number of relationships per concept (AVG_{rel}):

$$S_A(H)=1-\frac{1}{AVG_{rel}}=1-\frac{\sum_{i=0}^{m} S(i,H)}{\sum_{i=0}^{m} i*S(i,H)}$$

Where *m* is the highest level in H. The interpretation of $S_A(H)$ is as follows: if the average number of relationships per concept decreases, then $S_A(H)$ also decreases and it implies a simplification of the structure; otherwise, it implies an increase of the structural complexity. Hence, when the structural complexity function goes through a downward weight-shifting, $S_A(H)$ also decreases.

5.2.2 Accumulated Structural Complexity Measure

The structural complexity measure is still not completely justifying since it fails to reflect the impact of concepts' migrations from one level to another when the hierarchy is transformed from one state to the next. For instance, there was a large increase in the number of concepts on Level 1 for the July 2007 SNOMED. A reason for this phenomenon was a "weight shifting" from the higher-numbered Level 3 towards the lower-numbered Level 1.The condition where one element of (S(0), S(1),…, S(m)) increases while another element decreases does not communicate the decrease in structural complexity that took place. To reflect the above described downward weight-shifting phenomenon, an accumulated structural complexity measure is desirable, which not only measures the changes in the number of concepts at the different levels but also reflects the direction of the migration.

One would like to define a structural complexity measure that will enable a comparison of two states of the same SNOMED hierarchy and express the situation where a hierarchy in one state is more complex than in another.

Consider, for example, a downward weight-shifting transformation that occurs when, say, 20 concepts on Level 2 in a hierarchy H at time t (denoted H_t) have lost one relationship at the state $t+1$. In such a case, the total number of concepts in H_t and in H_{t+1} is equal, and one would consider H_{t+1} to be structurally less complex. However, the structural complexity function S does not express this fact. To illustrate this, assume that H_t has five levels with 50 concepts each. Then the S sequence for H_t is (50, 50, 50, 50, 50), and it is (50, 70, 30, 50, 50) for H_{t+1}. By comparing these sequences, it is not possible to judge which is more complex since the $S(1, H_{t+1}) > S(1, H_t)$, but $S(2, H_{t+1}) < S(2, H_t)$ (while all other components are equal).

To achieve the purpose of defining a structural complexity measure that can quantify that $H_{t+1}$is less complex than H_t, the Accumulated Structural Complexity Measure function S_c from S is defined as follows:

$S_c(0, H) = S(0, H)$;

And for $j = 1, \ldots, m$

$$S_c(j, H) = \sum_{i=0}^{j} S(i,H)$$

In sequence notation with respective to all levels, one can get (50, 100, 150, 200, 250) for H_t and (50, 120, 150, 200, 250) for H_{t+1} from the function S_c. In this case, $S_c(1, H_{t+1}) > S_c(1, H_t)$, while all other components are equal.

In general, for two hierarchy states H_t and H_{t+1} of the same total number of concepts, and with m levels, one can say H_{t+1} *dominates* H_t if

(1) There exists p ($0 \leq p < m$) such that $\forall i, p \leq i \leq m$, $S_c(i, H_{t+1}) \geq S_c(i, H_t)$.

and

(2) There exist j and k ($p \leq j < k < m$) such that $\forall i, j \leq i \leq k$, $S_c(i, H_{t+1}) > S_c(i, H_t)$.

According to condition (1), $S_c(i, H_{t+1}) \geq S_c(i, H_t)$ is only required beyond the Level p. Condition (2) states that there exists an interval of Levels $k - j$ above p reflecting an overall downward weight-shifting transformation from H_t to H_{t+1}.

When H_{t+1} dominates H_t, one can say that the hierarchy state H_t is structurally more complex than the hierarchy state H_{t+1}. Such a transformation may involve a simple downward weight-shifting between two consecutive levels, as in the example above, or it may involve more complex transformations. For example, some concepts in Level 2 lose one relationship while less concepts in Level 1 gain one relationship, so that the net change is a downward weight-shifting. Other more complex transformations may involve more than two levels, e.g., a net downward weight-shifting from Level 2 to Level 1, a net downward weight-shifting from Level 3 to Level 1, and a net downward weight-shifting from Level 3 to Level 2. For such a combination of two or three downward weight-shifting, there will be an interval [1, 2] of indices such that $S_c(i, H_{t+1}) > S_c(i, H_t)$, for $1 \leq i \leq 2$. (Note that in such a case, $S_c(3, H_t) = S_c(3, H_{t+1})$, due to the accumulative nature of S_c.)

Now let us illustrate the domination between two actual states of the Specimen Hierarchy for the July 2004 release and the July 2007 release of SNOMED. Table 5.1 shows the structural complexity function S for the Specimen Hierarchy of July 2004 and July 2007, a duration when the auditing effort [2, 70] took place, reviewed by K. Spackman and implemented in SNOMED. Comparing the values, one can see that 2007 is larger for Levels 1 and 2, while 2004 is larger for Levels 0, 3, and 4. Thus, one cannot conclude which hierarchy state is more complex.

Table 5.2 shows the similar comparison for S_c. Here, one can see a clear domination of the hierarchy for 2007 over 2004, implying that the Specimen hierarchy of 2007 is structurally simpler. Hence, in this case, the auditing effort helped to turn the Specimen hierarchy into a structurally simpler hierarchy.

Table 5.1 Number of Concepts for Levels (2004 vs. 2007)

Level(*l*)	**# Concepts (2004)**	**#Concepts (2007)**
	S(*l*)	**S(*l*)**
0	29	21
1	399	468
2	430	517
3	194	48
4	4	2
Total:	**1,056**	**1,056**

Table 5.2 Cumulative Number of Concepts for Levels (2004 vs. 2007)

Level(l)	**# Concepts (2004)**	**#Concepts (2007)**
	$S_c(l)$	**$S_c(l)$**
0	29	21
1	428	489
2	858	1,006
3	1,052	1,054
4	1,056	1,056

It is noted that in case H_{t+1} dominates H_t, H_{t+1}will also have a lower aggregated structural complexity function since the denominator $\sum_{i=0}^{m} i * S(i,H)$ is decreased in H_{t+1}while the numerator of the fraction in the formula did not change. For example, $S_A(H)$ for the Specimen hierarchy was decreased from 1 – 1056/1827 = 0.422 in 2004 to 1 – 1056/1654 = 0.362 in 2007.

The conditions (1) and (2) are given for the case where the total number of concepts in H_{t+1} is equal to that in H_t. In case the number of concepts in H_{t+1} is smaller or larger than in H_t, a scaling will be needed to bring the number of concepts in line to enable a comparison.

For the scaling, one can look at the percentage of the number of concepts in each level. The scaling is illustrated with the July 2008 release of 1,173 concepts, to be compared with the July 2007 release of 1,056 concepts. Table 5.3 shows the computation involved in the scaling.

Table 5.3 Scaling for the 2008 Specimen Hierarchy

Level	**# in Level**	**% of Level**	**Proportional level reduction**	**Scaled #**
0	20	2	2	18
1	397	34	40	357
2	450	38	45	405
3	293	25	29	264
4	13	1	1	12
Total	**1,173**	**100**	**117**	**1,056**

The percentage of the levels appears in column 3. The level difference (up or down) between the number of concepts in the two hierarchy states is distributed between the levels according to their percentages. Column 4 shows the proportional distribution of the 117 (= 1173 – 1056) concepts among the levels. The number of concepts in the new hierarchy state is modified (up or down) according to the level differences to yield a distribution of the number of concepts in hierarchy state H_t, according to the level percentages of hierarchy state H_{t+1}. The last column of Table 5.3 shows the scaled level numbers obtained in reducing the size of H_{t+1} (1,173) into the size of H_t (1,056). For example, the number in Level 1, 357, is 34% of the 1,056 scaled down size, rather than the 397 actual number of concepts in Level 1 of July 2008, which is 34% of the total of 1,173 concepts. The scaling enables a fair comparison of the cumulative structural complexity functions of two hierarchy states of different sizes to check for possible domination.

5.2.3 Hierarchical Complexity Measures

Another complexity measure concentrates on what is happening inside an area. An area may have several roots. Those roots are semantically independent of one another since none sits in an ancestor/descendant relationship to any other. Each root defines a partial-area, named after it, containing all concepts that are its specializations in the area. Each partial-area expresses an overarching semantics for its constituent concepts: each being a kind of the root concept. For example, all 19 concepts in the partial-area *Device specimen* in the {*identity*} area (Figure 3) are concepts that are specimens derived from various devices, such as *Catheter specimen*. That is, the division of an area into partial-areas serves to divide all concepts of the same structure (expressed by the area's name) into groups of semantically similar concepts. The semantics of each group is captured explicitly by the partial-area's name. Thus, one can make an assumption that a set of the same number of concepts with the same structure having fewer sub-hierarchies is simpler than one with more sub-hierarchies. In the context of the partial-area taxonomy, an area with fewer partial-areas is simpler than an area with more partial-areas (assuming the same number of concepts), as it contains a smaller variety of concepts. Similarly, an area with more concepts, but with the same number of partial-areas as before, is considered simpler. The above assumption is called the *hierarchical assumption.*

The ratio of the number of partial-areas to the number of concepts in an area can be used as a good measure of the hierarchical complexity of the area. In fact, the hierarchical complexity ratio of an area X is defined as follows. Let $P(X)$ be the set of partial-areas in X and $E(X)$ denote the extent of X(i.e., set of concepts of X). Then the Hierarchical Complexity (HC) Ratio $HC(X) = \frac{|P(X)|}{|E(X)|}$, where $|P(X)|$ ($|E(X)|$) is the number of partial-areas (concepts) in X (i.e., the cardinality of $P(X)$ ($E(X)$)). The idea behind the formula is that the semantics of each group is captured explicitly by the partial-area's name. Thus, an area with fewer partial-areas for the same number of concepts is an area with a smaller number of sets with different semantics. So, if the hierarchical complexity ratio gets smaller, then the hierarchical structure gets simpler. For example, in SNOMED 2004, the area {*substance*} had 56 concepts distributed across 15 partial-areas, for a hierarchical complexity ratio HC({*substance*}) = $\frac{15}{56}$ = 0.27. In SNOMED 2007, the same area has 81 concepts in ten partial-areas, for a ratio of 0.12. Hence, this area became simpler between 2004 and 2007.

5.2.4 Aggregated Hierarchical Complexity Measure

One can calculate the hierarchical complexity ratio for a whole level of areas having the same number of relationships. The aggregated hierarchical complexity ratio HC(i, H) for a Level i ($0 \leq i \leq m$) of the hierarchy H of m levels is defined by the following formula:

$$HC(i,H) = \frac{\sum_{X \in A(i,H)} |P(X)|}{\sum_{X \in A(i,H)} |E(X)|}$$

Where A(i, H) is the set of areas in the i^{th} level in hierarchy H. For example, there were 399 concepts in 153 partial-areas exhibiting exactly one relationship in 2004. Hence, the hierarchical complexity ratio HC(1, H_{2004}) for Level 1, obtained by dividing the number of its partial-areas, 153 into the number of 399 concepts for Level 1, is 0.38. The ratio for the same level for 2007 is HC(1, H_{2007})=$\frac{45}{468} = 0.10$. Therefore, as a whole, the aggregated hierarchical complexity of all areas of one relationship for the Specimen hierarchy became simpler from its state H_{2004} to its state H_{2007}. The hierarchical complexity for all levels will be compared.

This measure can also be applied for the whole hierarchy H of m levels.

$$\text{HC(H)} = \sum_{i=0}^{m} HC(i,H) = \frac{\sum_{i=0}^{m} \sum_{X \in L(i)} |P(X)|}{\sum_{i=0}^{m} \sum_{X \in L(i)} |E(X)|}$$

A possible impact of auditing is discovering that the root of a small partial-area, especially a singleton (i.e., a one-concept partial-area), should be a child of a concept in another partial-area. Hence, the small partial-area will be absorbed into the new parent's partial-area. For example, in 2004, {*morphology*} had nine partial-areas, four of which were: *Specimen from abscess*, *Specimen from ulcer*, *Specimen from wound*, and *Lesion sample*. In 2007, the first three of these became part of the expanded *Lesion sample*, which then consisted of all 14 concepts in {*morphology*}. HC({*morphology*}) went from $\frac{9}{15} = 0.6$ to $\frac{1}{14} = 0.07$, reflecting a hierarchical simplification that this area underwent when it was realized that abscess, ulcer, and wound were all kinds of lesions.

For the hierarchical measure, the complexity functions for an area, a level, and a whole hierarchy are defined. Both kinds of measures will be utilized to explore the impact of the application of an auditing process [2] on the complexity measure by examining the impact on the Specimen hierarchy.

An auditing process was applied again to this hierarchy in its 2007 release. An in-depth audit of all 255 singletons is performed. The larger partial-areas only had their names reviewed. This review will enable the identification of possible duplicates or missing IS-A relationships from the roots of the singletons to other partial-areas. While concentrating on the singletons, the auditors watched for propagation of errors to other partial-areas, and pursue these if appropriate.

It is hypothesized that in spite of the previous audits that were performed on the Specimen hierarchy for the 2004 release; there still were errors to be found. As a matter of fact, correction of previous errors may have caused new errors to arise or lead to the exposure of errors that remained hidden before. Another audit effort is also conducted on overlapping concepts of the partial areas in the Specimen hierarchy of July 2007 SNOMED release. This study is reported in [59]. Finally, the impact of this extra auditing effort applied to the Specimen hierarchy is investigated on the complexity measures that have postulated. An interesting question is whether the impact of the two audit efforts on the complexity measures is similar.

5.3 Results

First, let us apply the complexity measures to compare the state of two versions of the Specimen hierarchy over a long period of time, irrespective of any auditing. In particular, let us compare the version of July 2004 with that of January 2011. During that seven-year interval, the number of concepts grew from 1,056 to 1,329, while the number of relationships grew from 1,857 to 2,553 (with the average number of relationships per concept growing from 1.75 to 1.92). This growth is reflected by the aggregated structural hierarchy, which increased from 0.422 to 0.470.

Table 5.4 Scaling for the 2011.01.31 Specimen Hierarchy

Level	# in Level	% of level	Proportional level reduction	Scaled #
0	28	2	6	22
1	433	33	90	343
2	522	39	106	416
3	334	25	68	266
4	12	1	3	9
Total	**1,329**		**273**	**1,056**

To obtain the more detailed picture about what happened at the various levels, it helps to compare the structural complexity measure and the accumulated structural complexity measure for the two releases. (As discussed, scaling down is used for the Jan. 2011 release due to its greater number of concepts (see Table 5.4)). The values of the structural complexity and the accumulated structural complexity are given in Tables 5.5 and 5.6, respectively. In Level 1, there were 343 concepts (after scaling) in January 2011 in comparison to 399 in July 2004. As shown in Table 5.6, the accumulated structural complexity values for Level 1 were 365 and 428, respectively. The main change in the structural complexity is due to the growth of Level 3 at the expense of Level 1. When considering the absolute number of relationships, growth occurred in all levels except Level 0, but it was highest for Level 3 and also meaningful for Level 2. From Table 5.6, one can see that H_{2004} dominates H_{2011} for all levels. Hence, H_{2011} is structurally more complex.

Table 5.5 Structural Complexity Measures (2004, 2011)

Level(*i*)	**$S(i, H_{2004})$**	**$S(i, H_{2011})$**
0	29	22
1	399	343
2	430	416
3	194	266
4	4	9
Total:	**1,056**	**1,056**

Table 5.6 Accumulated Structural Complexity Measures (2004, 2011)

Level(*i*)	**$S_C(i, H_{2004})$**	**$S_C(i, H_{2011})$**
0	29	22
1	428	365
2	858	781
3	1,052	1,047
4	1,056	1,056

The hierarchical complexity measures HC for both releases appear in Table 5.7. In total, one can find values of 0.43 for H_{2004} and 0.31 for H_{2011}, making the latter less hierarchically complex, with less and larger partial areas and thus more cohesiveness. Looking at values for the various levels, one can see that the decrease in the hierarchical complexity measure occurs in all levels except Level 0, which is mainly due to the dramatic decrease in Level 1 caused by the large decline in the number of partial areas with one relationship.

Table 5.7 Hierarchical Complexity Measures (2004, 2011)

Level	**2004**			**2011**		
(*i*)	***|P(i)|***	***|E(i)|***	***HC(i)***	***|P(i)|***	***|E(i)|***	***HC(i)***
0	1	29	0.03	1	28	0.04
1	153	399	0.38	46	433	0.11
2	186	430	0.43	196	522	0.38
3	107	194	0.55	155	334	0.46
4	4	4	1.00	11	12	0.92
Total	451	1,056	0.43	409	1,329	0.31

Two audits were conducted on SNOMED's Specimen hierarchy. In the first, various auditing techniques were applied to the July 2004 release. The techniques and results were documented in [2, 70]. The audit reports were submitted to Dr. K. Spackman, currently Chief Terminologist of IHTSDO, an international organization in charge of developing and distributing SNOMED. The errors approved by Dr. Spackman were forwarded to the College of American Pathologists (CAP) for correction in SNOMED. The July 2007 release reflects the correction of these errors.

The second audit, comprising three separate auditing efforts, took place on the 2007 Specimen hierarchy. During the first effort, all partial-areas of one concept (singletons) were reviewed. The report on this effort appears in the current research. In the second and third efforts, all overlapping concepts of partial-areas and a set of non-overlapping concepts of a control sample, respectively, were reviewed. These efforts were reported in [63]. As with the audit on the 2004 version, corrections of errors which were found in all three efforts on the 2007 version and approved by Dr. Spackman, were implemented in SNOMED. The corrections are reflected in the July 2008 release.

To assess the impact of the first auditing effort on the complexity of the Specimen hierarchy, the complexity measures for 2004 and 2007 will be compared. Similarly to assess the impact of the second auditing effort, the complexity measures for 2007 and

2008 will be compared. For convenience, one will refer to the states of the Specimen hierarchy as H_1 for 2004, H_2 for 2007, and H_3 for 2008.

5.3.1 Structural Complexity Measures

Table 5.4 compares the number of concepts of all different levels for the three states of the Specimen hierarchy. For example, on Level 1 and Level 2, the values of the structural complexity function $S(1, H_{2007})$ and $S(2, H_{2007})$ reflect a large increase for concepts with one and two relationships in 2007, representing many more concepts with lower structural complexity as compared to 2004. The increase for Levels 1 and 2 is from 399 and 430 in 2004 to 468 and 517 in 2007, respectively. These increases are balanced by the decrease in concepts on Level 3 from 194 to 48. The total number of concepts of H_1 and H_2 is the same, following the initial decrease and subsequent increase due to changes in intermediate states as reported earlier. So the total number of concepts of H_1 and H_2 is equal by coincidence.

Interestingly, the picture is reversed when comparing H_2 and H_3. A large decrease occurs for Levels 1 and 2, balanced by an increase in Levels 3 and 4. Note that the number of concepts in the Specimen hierarchy in 2008 was actually 1,173, and the 1,056 total listed for H_3 reflects the scaling operation to enable a fair comparison of structural complexity (see Section 5.2).

It is noted that the decreases in $S(1, H_3)$ and $S(2, H_3)$ in 2008 from the corresponding numbers in 2007, are not as sharp as is seems from Table 5.8. Table 5.8 shows the scaled down numbers. The actual $S(1, H_3) = 397$ and $S(2, H_3) = 450$ (see Table 5.3) still reflect a decrease versus H_2 but are in line with $S(1, H_1) = 399$ and $S(2, H_1) = 430$.

Table 5.8 Structural Complexity Measures S(*i*, H) for 2004, 2007, 2008

Level(*i*)	**S(*i*, H_{2004})**	**S(*i*, H_{2007})**	**S(*i*, H_{2008})**
0	29	21	18
1	399	468	357
2	430	517	405
3	194	48	264
4	4	2	12
Total:	**1,056**	**1,056**	**1,056**

5.3.2 Accumulated Structural Complexity Measures

Table 5.5 shows the accumulated structural complexity measures S_C for H_1, H_2, and H_3. As already shown in Section 5.2 above, H_2 dominates H_1, implying that H_2 is a less structurally complex hierarchy state. On the other side, H_2 also dominates H_3. Hence, H_3 is a more complex hierarchy state than H_2. When comparing H_1 and H_3, one can see that H_1 dominates and is thus less complex than H_3. Hence, H_3 is the structurally most complex hierarchy state of these three states for the Specimen hierarchy.

Table 5.9 Accumulated Structural Complexity Measures for 2004, 2007, 2008

Level(*i*)	**S_C(i, H_{2004})**	**S_C(i, H_{2007})**	**S_C(i, H_{2008})**
0	29	21	19
1	428	489	378
2	858	1,006	755
3	1,052	1,054	1,046
4	1,056	1,056	1,056

The aggregated structural complexity measure for H_1, H_2, and H_3 is 0.422, 0.362, and 0.474, respectively.

5.3.3 Hierarchical Complexity Measures

Table 5.6 presents in detail the dramatic hierarchical simplification of the areas on Level 1 from 2004 to 2007. This can be seen by examining the ratio $|P(X)|/|E(X)|$ of each area, separately. The large decrease in the structural complexity occurs mainly due to the large decrease in the number of partial-areas for the areas {*morphology*}, {*topography*}, and {*identity*}. For example, in{*topography*}, the ratio decreased from 0.38 to 0.08. In contrast, little change is seen from 2007 to 2008.

Table 5.10 The Hierarchical Complexity for Areas with One Relationship (2004, 2007, 2008)

Area (X)	**2004**			**2007**			**2008**		
	|*P(X)*|	**|*E(X)*|**	***HC(X)***	**|*P(X)*|**	**|*E(X)*|**	***HC(X)***	**|*P(X)*|**	**|*E(X)*|**	***HC(X)***
substance	15	56	0.27	10	81	0.12	10	98	0.10
morphology	9	15	0.60	1	14	0.07	1	13	0.08
topography	112	297	0.38	25	333	0.08	23	245	0.09
procedure	9	12	0.75	7	20	0.35	8	20	0.40
identity	8	19	0.42	2	20	0.10	2	21	0.10

The hierarchical complexity ratio in Table 5.11 is given with respect to the taxonomy’s various levels for the three hierarchy states. The whole hierarchy became simpler from 2004 to 2007 with a decrease in the number of partial-areas from 451 to 361, as seen in the last row. Level1 became much simpler because there are more concepts and quite a bit fewer partial-areas, and thus a much lower hierarchical complexity. Level2 became somewhat more complex hierarchically from 2004 to 2007, as there are more concepts, but relatively even more partial-areas. On Level 3, there are fewer concepts, even fewer partial-areas, and thus higher hierarchical complexity. Comparing hierarchical complexity for 2007 and 2008, one can see a further small decrease due to the increase of the number of concepts for H_3 while the number of partial areas increased only slightly. The major contributions for the decrease come from the sharp decrease in |*P(2)*| and sharp increase in |*E(3)*|.

Table 5.11 The Hierarchical Complexity Measures for Levels (2004, 2007, 2008)

Level	**2004**			**2007**			**2008**		
(*i*)	***|P(i)|***	***|E(i)|***	***HC(i)***	***|P(i)|***	***|E(i)|***	***HC(i)***	***|P(i)|***	***|E(i)|***	***HC(i)***
0	1	29	0.03	1	21	0.05	1	20	0.05
1	153	399	0.38	45	468	0.10	44	397	0.11
2	186	430	0.43	269	517	0.52	178	450	0.40
3	107	194	0.55	44	48	0.92	133	293	0.45
4	4	4	1.00	2	2	1.00	12	13	0.92
Total	451	1,056	0.43	361	1,056	0.34	368	1,173	0.31

In the auditing results for the 255 singletons in 2007, it is found that errors of different kinds. A sample of such errors is shown in Table 5.12. For example, it was discovered that *Edema fluid sample* in the area {*morphology*, *substance*} has a parent *Fluid sample* that is incorrect because it is too general. The parent should instead be *Body fluid sample*. The auditing was performed by JX who is a MD with experience using SNOMED in practice and research. The errors were reviewed by Dr. Spackman. Only errors confirmed by him are reported in Tables 5.8 and 5.9.

Table 5.12 Sample of Errors Discovered in the Specimen Hierarchy

Concept Name	Area	Error Type	Correction
Edema fluid sample	morphology, substance	Incorrect parent: Fluid sample	Correct parent: Body fluid sample
Vein sample	topography, substance	Incorrect parent: Specimen from heart	Correct parent: Cardiovascular sample
Specimen from pleura obtained by fine needle aspiration procedure	topography, procedure	Missing parent	Add parent: Specimen obtained by aspiration
Specimen from thoracic mesothelium obtained by open thoracotomy	topography, procedure	Missing parent	Add parent: Surgical excision sample
Specimen from breast obtained by image guided core biopsy	topography, identity, procedure	Incorrect identity relationship to Core biopsy needle	Delete relationship
Drainage fluid sample	morphology, procedure, substance	Incorrect morphology relationship to Discharge	Delete relationship
Tissue specimen obtained from anus by polypectomy	topography, procedure	Missing substance relationship	Add substance relationship to Body tissue material
Buccal smear sample	topography, procedure	Incorrect procedure target: Biopsy	Correct target: Smear
Tissue cell sample	topography, procedure	Ambiguous concept	Retire it
Cervical secretion sample	topography, substance	Duplicate concept with Cervical mucus specimen	Remove duplicate concept. Combine parents
Skin ulcer swab	topography, morphology, procedure	Duplicate procedure targets: Taking skin swab, Swabbing skin area	Consolidate the two targets

In Table 5.13, the 88 confirmed errors are divided according to their kind. The majority of them——69 in total—were incorrect or missing parents. These errors were corrected in the 2008 release. The next two errors in terms of frequency of occurrence, namely, missing relationships and incorrect relationships, occurred only for eight and four concepts, respectively. In Section 5.4, the kinds of errors are related to the findings regarding the complexity measures. For this reason, one can include here also the distribution of the kinds of errors found for the other auditing efforts with respect to the 2007 Specimen hierarchy that involved 162 "overlapping" concepts and for a control sample of 85 concepts (see Tables 5.10 and 5.11, respectively). As reported in [63], the auditing for this project was performed by three auditors: GE, JX, and YC, all trained in medicine and medical terminologies, and experienced in auditing medical terminologies. For the overlapping concepts, there were a total of 53 errors of missing or incorrect children and a total of 74 missing or incorrect parents. There were 21 missing relationships. For the errors in the control sample, there was a total of 45 errors of missing or incorrect parents, eight errors of missing or incorrect children, and 55 cases of missing relationships.

Table 5.13 Kinds of Errors and Their Counts (Auditing Singletons)

Kind of Error	**#**
Incorrect parent	55
Missing parent	14
Missing relationship	8
Incorrect relationship	4
Incorrect target	3
Ambiguous concept	2
Duplicate concept	1
Duplicate targets	1
Total	**88**

Table 5.14 Kinds of Errors and Their Counts (Auditing Overlapping Concepts)

Kind of Error	#
Ambiguous concept	1
Missing child	48
Incorrect child	5
Missing parent	30
Incorrect parent	44
Missing relationship	21
Missing sibling	4
Incorrect target of relationship	5
Total	**158**

Table 5.15 Kinds of Errors and Their Counts (Auditing Non-overlapping Concepts (Control Sample))

Kind of Error	#
Missing parent	23
Incorrect parent	22
Missing child	6
Incorrect child	2
Missing relationship	55
Incorrect relationship target	2
Other types	6
Total	116

5.4 Discussion

In this research, one can set out to define complexity measures for a SNOMED hierarchy and explore the changes in those measures as the hierarchy goes through stages of auditing. Two kinds of measures are introduced. The first relates to the structure (set of relationships) defined for a hierarchy. This measure is closely related to the area taxonomy that have previously introduced [2]. The second kind of complexity measure depends on the ratio of the partial-areas to the number of concepts. This kind is defined in the context of the partial-area taxonomy [2] and seeks to measure the cohesiveness of the hierarchy.

When auditing, one can eliminate or change incorrect knowledge elements and add missing ones. The idea of a connection of some sort between auditing and complexity stems from the possibility that errors in the modeling of concepts cause some disorderliness in the knowledge of a hierarchy. If so, the auditing may help to decrease disorderliness. If disorderliness is expressed by an increase in complexity of a hierarchy, then perhaps auditing will be manifested as a decrease of the complexity of the hierarchy.

However, one needs to be aware that complexity also relates to how extensive and involved the knowledge represented in the hierarchy is, and not necessarily errors. Hence, the connection between auditing and complexity may be subtle, depending on the kind of auditing applied, and also on any further development that has taken place in a hierarchy. Furthermore, there may be differences between an initial audit phase and a subsequent audit phase.

As is seen in the Section 5.3, there is a difference in the changes between the two auditing periods tracked in this study. The first audit phase yields a decrease in complexity measures of both kinds. First, let us concentrate on the structural complexity measures. The aggregated structural complexity was reduced from 0.422 to 0.362 reflecting a reduction of 203 relationships (from 1,857 to 1,654) between 2004 and 2007. (This count does not include occurrences of multiple targets for the same relationship with respect to the same source concept, which are not reflected in the definition of the structural complexity.) The reduction of 203 erroneous relationships in a hierarchy of 1,056 concepts is a meaningful improvement in both quality and simplicity. The amount of incorrect relationships is even higher than it seems to be if one also considers the relationships that were found to be missing and were subsequently added (e.g., for eight

(= 29–21) of the area Ø on Level 0), since those cancel the impact of the same number of deleted relationships. Obviously, it is imperative that concepts have the correct relationships, even if it makes them more complex. To illustrate such an example, in 2004, the partial-area *Specimen from digestive system* had an extraneous *identity* relationship that was subsequently removed from its 38 concepts [2]. This improvement in structural complexity obtained by the movement of concepts from Levels 3 and 4 to Levels 1 and 2 (see Table 5.8) is properly captured by the accumulated structural complexity measure for which H_2 dominates H_1 (see Table 5.9).

The change from 2004 to 2007 in the hierarchical complexity (HC) measure is more involved. Globally, the ratio of HC decreased from 0.43 to 0.34 due to the large decrease of 90 (20% of 451) in the number of partial-areas. But in the levels only, HC(1) decreased, while HC(0), HC(2), and HC(3) increased. There were several different factors at play here. The main contribution was the finding that many roots of small partial-areas (mainly singletons) were missing IS-A relationships to concepts in other areas. When those IS-As were added, these concepts ceased to be roots of independent partial-areas, and all concepts of their partial-areas became part of another partial-area. In general, these corrections turned the hierarchy into a more cohesive structure with fewer and larger partial-areas. For example, the nine partial-areas of {*morphology*} turned into one called *Lesion sample*, consisting of 14 concepts. This consistent and dramatic transformation in all areas of Level 1 is documented in Table 5.10.

Another phenomenon was the increase in the number of partial-areas in Level 2 from 186 (2004) to 269 (2007). This increase was much larger than the increase in the number of concepts: 430 to 517. This phenomenon centers around two areas, {*topography*, *substance*} of 90 concepts and {*topography*, *procedure*} of 380 concepts (see Figure 2). Another page and a half (omitted) were required to draw them. Those two areas have so many partial-areas, 31 and 194, respectively, that they cannot be displayed in Figure 3. Those partial-areas were created by the many combinations of ten partial-areas of {*substance*} and 25 partial-areas of {*topography*} (see Level 1 in Figure 3). It is found that a similar situation for the combinations of the partial-areas of {*topography*} with those of {*procedure*}. As it happened, many procedures can be applied to many body parts leading to all those 194 partial-areas of {*topography*, *procedure*}. Finally, at Level 3, the 48 concepts left in 2007, after the movement of 194 to lower levels, resided in as many as 44 partial-areas, yielding an HC = 0.92. However, when all these interplays were combined, the impact of the consolidation of partial-areas in Level 1 outweighs the others, and the result of the auditing phase for 2004 was a more cohesive Specimen hierarchy with lower hierarchical complexity.

Hence, as a result of the 2004 audit phase, the Specimen hierarchy became simpler both structurally and hierarchically. That is, in parallel to many errors being corrected, the hierarchy became more cohesive and its concepts less structurally complex. The average number of relationships per concept was reduced from 1.76 to 1.57.

The picture for the second audit phase applied for the 2007 Specimen hierarchy is very different. The structural hierarchy increased even beyond the original 2004 level. For example, the aggregated structural complexity grew to 0.474, 30% higher than in 2004. This increase is well reflected by the cumulative structural complexity measure, where H_3 is dominated by both H_1 and H_2, that is, H_3 is more complex than both. On the other hand, the hierarchical complexity continued to decrease a little in this auditing phase. The number of partial-areas increased minimally from 361 to 368, but the number of concepts increased by 11% to 1,173. Thus, HC increased, with a larger number of concepts arranged in about the same number of partial-areas, implying an increase in hierarchical cohesiveness.

One question which arises is: what is the reason for the difference between the two phases regarding the structural complexity? A second question is what is the reason for the difference in the direction of change for the two complexity measure during the second phase? To look for answers, one needs to look at the nature of the auditing techniques used in the second phase and the kinds of errors found, as documented in Tables 5.13, 5.14, and 5.15.

Altogether, in the three separate auditing efforts on 2007 Specimen hierarchy (502 concepts), a little less than half of the hierarchy was reviewed. According to the confirmed hypotheses of [62] and [59], those are concepts with the highest likelihood of being in error. The errors relevant for structural complexity are incorrect relationships, which should be removed, and missing relationships, which should be added. Only in the context of the singleton auditing did one find four errors of incorrect relationships. On the other hand, missing relationships were found during all three auditing efforts. The audits exposed eight missing relationships for the singletons, 21 for the overlapping concepts, and 55 for the non-overlapping control sample. Hence, a total 84 missing relationships and four incorrect relationships were reported, leading to a net gain of 80 relationships in the Specimen hierarchy. Obviously, such corrections will increase the structural complexity.

An obvious question is: what is the reason for such a difference in the change of the number of relationships between the first audit and the second audit. In 2004, the overlapping concepts were not audited but singletons and other small partial-areas were audited. It is discovered that so many incorrect relationships that after accounting for the added relationships, one still had a net decrease of 203 relationships.

One possibility is that incorrect relationships, that need to be deleted, are relatively easier to detect than missing relationships that need to be included. The erroneous knowledge asserted by incorrect relationships are perhaps stick out like a sore thumb. These relationships are detectable even on a cursory review. On the other hand, to detect a missing relationship, one has to absorb all the existing knowledge and then surmise that some is missing, which can be a much more demanding mental task. Hence, it is assumed that almost all incorrect relationships were discovered in the 2004 audit, while only a portion of the missing relationships were uncovered and added in that initial audit phase.

The errors relevant for the hierarchical complexity measure are missing parent relationships and perhaps incorrect parent relationships. Tables 5.13, 5.14, and 5.15 show respectively 14, 30, and 23 missing parent relationships in the corresponding auditing projects. The numbers of incorrect parent relationships are 55, 44 and 22, respectively. Thus it is found that a total of 67 missing parent relationships and 121 wrong parent relationships during the second audit phase of 2007.

As explained earlier, when missing parent relationships are added, it sometimes implies a partial-area merging into another partial-area. Based on the number of missing parent relationships found, one would expect to see a continuation of the reduction in the number of partial-areas similar to the impact of the previous audit of 2004. But what one can see is a small increase of seven partial-areas. Again, there may be an interplay of several factors here. For example, it is not clear what the impact of handling errors of incorrect parent relationships on the number of partial-areas is. A deletion of such a relationship may separate one partial-area into two, while a replacement by a parent relationship to another concept has typically no impact on the number of partial-areas.

Also, one can see that a net of 117 concepts were added to the Specimen hierarchy during this period, which are attributed to the routine developmental work done by CAP (which maintained SNOMED at the time). New partial-areas are likely to arise as a result of that development. On the other hand, one can see among these 117 concepts added the inclusion of general concepts like *Specimen from head and neck structure* or *Specimen from head and neck structure obtained by biopsy* that were added as parents of many existing concepts and some new concepts. This ended up turning many small-partial areas into one large one. In summary, the period of 2007–2008 was not a period of just auditing activity but one of combined development and auditing. Hence, it is difficult to isolate the impact of the auditing itself on the complexity. This is discussed further below.

Limitations and Future Work

In this work, it is shown that the Specimen hierarchy of SNOMED became simpler according to the two complexity measures due to an initial auditing effort. The situation for a subsequent auditing effort was a mixed report due to several issues discussed above. More experiments with other hierarchies of SNOMED or similar terminologies (e.g., NCIt) are needed to further study the connection between complexity and the impact of auditing on a hierarchy of a DL-based terminology [61, 79, 80] such as SNOMED.

A major limitation of the suggested complexity measures is that they are not applicable for SNOMED hierarchies without outgoing relationships. Eleven out of the 19 hierarchies (e.g., the Physical Object hierarchy) fall into this category. The concepts of these hierarchies just serve as targets for the relationships from other hierarchies. Thus, the area taxonomy and the partial-area taxonomy are not defined for them. (In [60], the converse abstraction network is introduced to handle such hierarchies.) Hence, neither of the two complexity measures is applicable. It is a research problem to identify what aspects of such a hierarchy need to be reflected in a complexity measure.

A particular problem one encountered was in reporting the complexity measure for the period 2007–2008, during which time the Specimen hierarchy went through three separate audits and, evidently, regular content development performed by CAP. It is noted that potentially, these two activities may influence changes in the complexity measures in different ways. A future research problem is to investigate the impact of content development on complexity measures of a hierarchy. One expects a different impact from that for auditing due to several factors. One also expects different impacts in early development stages versus later stages. For example, in early stages, new concepts are expected to be less complex (with few relationships), while in later stages, many simple concepts are already in the hierarchy and typically more complex concepts (with relatively more relationships) are added. Also, in earlier stages, many new concepts are expected to be roots of new partial-areas, while in later stages concepts are mostly joining existing partial-areas. Such phenomena will influence the complexity measures differently.

To investigate such phenomena, one has to identify precisely when a hierarchy goes through the different processes. For example, in the SNOMED releases of July 2008, July 2009, July 2010, and January 2011, it is observed that the following numbers of concepts in the Specimen hierarchy, respectively: 1,173, 1,236, 1,266, and 1,329. To ones' knowledge, there was no auditing activity for this hierarchy during this period. Hence, the period from July 2008 through January 2011 represented a time of just content development, for which one could investigate the impact on the complexity measures.

Another interesting research problem is what will happen if an audit will target the new concepts that were added to the Specimen hierarchy during a phase of content development, as for this period of July 2008 – January 2011. Would one see a decrease in the complexity due to such an audit, as one saw for the initial audit of 2004?

It is observed that an increase in the structural complexity and a decrease in the hierarchical complexity when comparing the releases of 2011 to that of 2004. This is not unexpected because concepts added later in the hierarchy's life cycle tend to be more complex and have more relationships, since the simpler ones in the lower-numbered levels already exist. The cohesiveness of the hierarchy tends to improve over time since concepts added later, more often join existing partial-areas than establish new ones.

However, this broad range comparison overlooks important facts that were exposed when tracking the impact of auditing efforts. For example, the decrease in the number of partial-areas in Level 1, causing the decrease in the hierarchical complexity, is due to the initial audit carried out with respect to the 2004 version. When comparing the hierarchical complexity of 2008 and 2011, no change is seen in spite of the addition of 156 concepts. Similarly, the structural complexity declined at first due to the initial 2004 audit, but later increased. These observations support ones opinion that more refined analysis is needed to reflect different change patterns for auditing versus content development. Furthermore, there are differences between changes occurring during the initial periods of a hierarchy's lift cycle and later periods when the hierarchy is in a more mature state. Such differences exist for both auditing activity and content development activity.

It is noted that the results of the 2004 auditing effort were visible in the July 2005 release, where the Specimen hierarchy had decreased in size to just 1,042 concepts. In subsequent releases, it grew back to 1,049 (July 2006), 1,052 (January 2007), and eventually to 1,056 (July 2007), its original size in July 2004.

One had the option of auditing the July 2005 release instead of the July 2007 release. During the July 2004 – July 2005 period, the Specimen hierarchy seemed to have gone strictly through auditing, as reported in [2, 70]. The following two years of release periods showed very slow growth of 12 concepts. It is not certain if these 12 concepts were added as a result of the auditing reports or some other auditing performed by CAP, or just reflected a slow development process. One decided to use the July 2007 release as both ending the first audit period (of 2004) and starting the second audit period (ending July 2008) for several reasons. First and foremost, the impact of the addition of the 12 concepts seems negligible compared to the major changes that resulted from auditing, as described in Section 5.3. Second, it was simpler to deal with only three states of the hierarchy, where H_2 represented both the end of the first period and the beginning of the second. Otherwise, one would have needed to process the July 2005 version, as the end state for the first audit period, and deal with the increase in size (from 1,044 concepts to 1,056). The third reason was that the coincidence of H_2 and H_1 having the same number of concepts enabled the direct comparison of the structural complexity measures of 2004 and 2007 without introducing scaling. This simplified the presentation of the results. Once such a comparison was done, it was easier to introduce the reader to scaling, necessary for the second period. It is also confident that the results for comparison with the July 2005 release would be very similar and no scientific gain would have been obtained for the price of a longer and more complex presentation. Note that no scaling was necessary for the hierarchical complexity measures, which are based on the number of partial-areas and the number of concepts (42).

5.5 Summary

Two kinds of measures were introduced to quantify the complexity of a SNOMED hierarchy. They are based, respectively, on characteristics of the area taxonomy and partial-area taxonomy abstraction networks that are previously introduced. Both networks are derived automatically via analysis of structural aspects of the hierarchy.

The structural complexity measure is proposed as a means to measure the lateral density of the hierarchy network by computing the ratio of the number of lateral relationships (non-hierarchical links) with respect to the number of concepts (nodes). The denser the network, the higher the aggregated structural complexity. The hierarchical complexity measure reflects the cohesiveness of the hierarchy by computing the ratio of semantically uniform concept groups (partial-areas) to the number of concepts. The fewer and larger the partial-areas, the lower the hierarchical complexity.

The two suggested measures offer a quantitative way to track a hierarchy over time and see the changes occurring in its density and cohesiveness. In particular, it is studied the changes occurring as a result of auditing efforts applied to the Specimen hierarchy. It is observed that during the initial audit period, both complexity measures decreased. The outcome of the subsequent audit period is mixed: the structural complexity increased, while the hierarchical complexity decreased.

CHAPTER 6

SUMMARY AND FUTURE WORK

This dissertation is aimed to develop automated or semi-automated methods that focus auditing of the SNOMED on those areas that are most likely to have errors. A medical terminology, such as SNOMED, is like a bridge that connects two sides – human cognitive processing and computer processing. However, previous research work indicates that the quality of SNOMED is often not up to the standard required for critical application. Therefore, quality assurance of SNOMED is extremely important for medical systems and EHR/EMR systems.

This research is built upon the previous work by extending and using previously defined abstraction networks for structural auditing of SNOMED. Converse relationships, derived from relationships in SNOMED's inferred view, have been used in the construction of a new kind of abstraction network, the Converse Abstraction Network (CAN), for a strict target hierarchy. An auditing methodology for such a hierarchy's incoming relationships based on the CAN has been presented. The results of applying this methodology to the Device sub-hierarchy indicate that the CAN is a useful auditing vehicle that can bring various aspects of the relationship structure to light and aid an auditor in refining and improving SNOMED.

The difference between Description Logic (DL) classification and the Abstraction Network (AN) methodology (mentioned in Section 4.2.2) are examined in identifying errors in SNOMED. DL classifiers can automatically identify logical inconsistencies in the terminology, while the AN methodology helps experts perform targeted manual reviews of the terminology by reducing its complexity and grouping the concepts by their structural and semantic properties. The differences between the two approaches were illustrated through two cases of errors identified in SNOMED.

Two complexity measures are explained in detail: one structural and one hierarchical. A global complexity measure is investigated for the structural complexity that is more detailed than just the total number of relationships and thus can measure the impact of the phenomenon of concepts moving between levels due to omitted or added relationships. They have been utilized to assess the hypothesis that following the application of an auditing process, the Specimen hierarchy got simpler. An auditing process has been further utilized to this hierarchy in its July 2007 release. Reflecting the typical situation of scarce resources for auditing, an in-depth audit of only the 255 singleton partial-areas has been done. The larger partial-areas only had their names reviewed. This review enabled the identification of possible duplicates or missing IS-A relationships from the (roots of) singletons to other partial-areas. While concentrating on the singletons, the auditor can watch for propagation of errors to other partial-areas, and pursue those if appropriate.

It is realized that in spite of the two previous audits performed on the Specimen hierarchy, there were still be errors to be found. As a matter of fact, correction of previous errors would cause new errors to arise or lead to the exposure of errors remaining hidden before. So, whether this extra auditing effort further simplifies the Specimen hierarchy according to the measures we have postulated has been checked.

Finally, the author proceeded from the assumption that “complex” concepts warrant particular attention in quality assurance activities pertaining to terminologies like

SNOMED. An auditing methodology is presented in which the author took such complex concepts to be those residing in special overlapping subsets of a SNOMED hierarchy defined with respect to the partial-area taxonomy. These so-called overlapping concepts in the Specimen hierarchy are identified programmatically and then put through a rigorous audit. Comparing these auditing results with results from a control set, a statistically significant of higher error rate among the overlapping concepts has been found.

In the future, the current study will be extended in the following directions: applying the abstraction networks (including area taxonomy and partial-area taxonomy) to the hierarchies with large numbers of concepts and rich sets of relationships; and associating the errors found in the source terminology, like SNOMED, with the UMLS semantic-type assignments and investigate the reasons for those errors. The overarching goal for the future research is to identify semantic-type assignment errors of SNOMED concepts in the UMLS by utilizing the previously defined partial-area taxonomy. The partial-area taxonomy's semantic division of a hierarchy's concepts will be used to facilitate the task of auditing semantic-type assignments. The SNOMED semantic divisions, as logic units, will be mapped to the UMLS version of SNOMED. Such semantically uniform groupings are more comprehensible and much easier for auditing UMLS semantic type assignments.

REFERENCES

[1] *IHTSDO: International Health Terminology Standards Development Organisation*. [cited March 30, 2011]; Available: www.ihtsdo.org

[2] Y. Wang, M. Halper, H. Min, Y. Perl, Y. Chen, and K. A. Spackman, Structural methodologies for auditing SNOMED, *J Biomed Inform,* vol. 40, pp. 561-81, Oct 2007.

[3] A. L. Rector, Clinical terminology: Why is it so hard?, *Methods of Information in Medicine,* vol. 38, pp. 239-252, 1999.

[4] J. J. Cimino, Review paper: coding systems in health care, *Methods Inf Med,* vol. 35, pp. 273-84, Dec 1996.

[5] J. J. Cimino, P. D. Clayton, G. Hripcsak, and S. B. Johnson, Knowledge-based approaches to the maintenance of a large controlled medical terminology, *J Am Med Inform Assoc,* vol. 1, pp. 35-50, Jan-Feb 1994.

[6] E. H. Shortliffe, G. O. Barnett, J. J. Cimino, R. A. Greenes, S. M. Huff, and V. L. Patel, Collaborative medical informatics research using the Internet and the World Wide Web, *Proceedings of AMIA Annual Symposium,* pp. 125-9, 1996.

[7] O. Bodenreider, The Unified Medical Language System (UMLS): Integrating biomedical terminology, *Nucleic Acids Research,* vol. 32, 2004.

[8] C. Lindberg, The Unified Medical Language System (UMLS) of the National Library of Medicine, *J Am Med Rec Assoc,* vol. 61, pp. 40-2, May 1990.

[9] O. Bodenreider, Using UMLS semantics for classification purposes, *Proc AMIA Symp,* pp. 86-90, 2000.

[10] *International Classification of Diseases. Ninth Revision, with Clinical Modifications (ICD-9-CM).* Ann Arbor, MI: Commission on Professional and Hospital Activities, 1978.

[11] M. Wood, The international classification of primary care--an update, *J Fam Pract,* vol. 28, pp. 147-9, Feb 1989.

[12] *Current Procedural Terminology*. Chicago, IL: American Medical Association, Continuously updated.

[13] *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association. Committee on Nomenclature and Statistics, 1994.

[14] W. Ceusters, Applying evolutionary terminology auditing to the Gene Ontology, *J Biomed Inform,* vol. 42, pp. 518-29, Jun 2009.

[15] J. Friedlin and C. J. McDonald, An evaluation of medical knowledge contained in Wikipedia and its use in the LOINC database, *J Am Med Inform Assoc,* vol. 17, pp. 283-7, May 1 2010.

[16] A. W. Forrey, C. J. McDonald, G. DeMoor, S. M. Huff, D. Leavelle, D. Leland, T. Fiers, L. Charles, B. Griffin, F. Stalling, A. Tullis, K. Hutchins, and J. Baenziger, Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results, *Clin Chem,* vol. 42, pp. 81-90, Jan 1996.

[17] *Logical Observation Identifiers Names and Codes (LOINC)*. [cited August 2011]; Available: http://loinc.org/

[18] *Medical Subject Headings* [cited March 30, 2011]; Available: http://www.nlm.nih.gov/pubs/factsheets/mesh.html

[19] C. Rosse and J. L. Mejino, Jr., A reference ontology for biomedical informatics: the Foundational Model of Anatomy, *J Biomed Inform,* vol. 36, pp. 478-500, Dec 2003.

[20] O. Bodenreider, Biomedical ontologies in action: role in knowledge management, data integration and decision support, *Yearb Med Inform,* pp. 67-79, 2008.

[21] W. R. Hersh and R. A. Greenes, SAPHIRE--an information retrieval system featuring concept matching, automatic indexing, probabilistic retrieval, and hierarchical relationships, *Comput Biomed Res,* vol. 23, pp. 410-25, Oct 1990.

[22] W. Hersh, D. H. Hickam, R. B. Haynes, and K. A. McKibbon, Evaluation of SAPHIRE: an automated approach to indexing and retrieving medical literature, *Proc Annu Symp Comput Appl Med Care,* pp. 808-12, 1991.

[23] W. R. Hersh and L. C. Donohoe, SAPHIRE International: a tool for cross-language information retrieval, *Proceedings of AMIA Annual Symposium,* pp. 673-7, 1998.

[24] V. R. Sutton and S. E. Hauser, Preliminary comparison of the Essie and PubMed search engines for answering clinical questions using MD on Tap, a PDA-based program for accessing biomedical literature, *Proceedings of AMIA Annual Symposium,* p. 1128, 2005.

[25] C. L. Cole, A. S. Kanter, M. Cummens, S. Vostinar, and F. Naeymi-Rad, Using a terminology server and consumer search phrases to help patients find physicians with particular expertise, *Stud Health Technol Inform,* vol. 107, pp. 492-6, 2004.

[26] C. W. Bishop, Alternate approaches to a UMLS, *Med Decis Making,* vol. 11, pp. S99-102, Oct-Dec 1991.

[27] S. T. Rosenbloom, R. A. Miller, K. B. Johnson, P. L. Elkin, and S. H. Brown, Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems, *Journal of the American Medical Informatics Association,* vol. 13, pp. 277-288, 2006.

[28] T. C. Rindflesch and M. Fiszman, The interaction of domain knowledge and linguistic structure in natural language processing: Interpreting hypernymic propositions in biomedical text, *Journal of Biomedical Informatics,* vol. 36, pp. 462-477, 2003.

[29] W. Ceusters, B. Smith, and L. Goldberg, A terminological and ontological analysis of the NCI Thesaurus, *Methods of Information in Medicine,* vol. 44, pp. 498-507, 2005.

[30] N. Sager, M. Lyman, N. T. Nhan, and L. J. Tick, Medical language processing: Applications to patient data representation and automatic encoding, *Methods of Information in Medicine,* vol. 34, pp. 140-146, 1995.

[31] *Protege*. [cited August 2011]; Available: http://protege.stanford.edu/

[32] N. F. Noy, N. H. Shah, P. L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D. L. Rubin, M. A. Storey, C. G. Chute, and M. A. Musen, BioPortal: ontologies and integrated data resources at the click of a mouse, *Nucleic Acids Res,* vol. 37, pp. W170-3, Jul 1 2009.

[33] N. F. Noy, M. Crubezy, R. W. Fergerson, H. Knublauch, S. W. Tu, J. Vendetti, and M. A. Musen, Protege-2000: an open-source ontology-development and knowledge-acquisition environment, *Proceedings of AMIA Annual Symposium,* p. 953, 2003.

[34] M. A. Musen, J. H. Gennari, H. Eriksson, S. W. Tu, and A. R. Puerta, PROTEGE-II: computer support for development of intelligent systems from libraries of components, *Medinfo,* vol. 8 Pt 1, pp. 766-70, 1995.

[35] M. A. Musen, H. Eriksson, J. H. Gennari, S. W. Tu, and A. R. Puerta, PROTEGE-II: a suite of tools for development of intelligent systems from reusable components, *Proc Annu Symp Comput Appl Med Care,* p. 1065, 1994.

[36] M. A. Musen, Dimensions of knowledge sharing and reuse, *Comput Biomed Res,* vol. 25, pp. 435-67, Oct 1992.

[37] *SNOMED International: The Systematized Nomenclature of Medicine Clinical Terms.* [cited July 30 2009]; Available: http://www.ihtsdo.org

[38] M. Q. Stearns, C. Price, K. A. Spackman, and A. Y. Wang, SNOMED clinical terms: overview of the development process and project status, *Proceedings of AMIA Annual Symposium,* pp. 662-6, 2001.

[39] K. Spackman, SNOMED RT and SNOMEDCT. Promise of an international clinical terminology, *MD Comput,* vol. 17, p. 29, Nov-Dec 2000.

[40] K. A. Spackman, K. E. Campbell, and R. A. Cote, SNOMED RT: a reference terminology for health care, *Proceedings of AMIA Annual Symposium,* pp. 640-4, 1997.

[41] H. Min, Y. Perl, Y. Chen, M. Halper, J. Geller, and Y. Wang, Auditing as part of the terminology design life cycle, *J Am Med Inform Assoc,* vol. 13, pp. 676-90, Nov-Dec 2006.

[42] J. R. Campbell, P. Carpenter, C. Sneiderman, S. Cohn, C. G. Chute, and J. Warren, Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures, *J Am Med Inform Assoc,* vol. 4, pp. 238-51, May-Jun 1997.

[43] C. G. Chute, S. P. Cohn, K. E. Campbell, D. E. Oliver, and J. R. Campbell, The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures, *J Am Med Inform Assoc,* vol. 3, pp. 224-33, May-Jun 1996.

[44] J. F. Penz, S. H. Brown, J. S. Carter, P. L. Elkin, V. N. Nguyen, S. A. Sims, and M. J. Lincoln, Evaluation of SNOMED coverage of Veterans Health Administration terms, *Stud Health Technol Inform,* vol. 107, pp. 540-4, 2004.

[45] J. J. Cimino, Auditing the Unified Medical Language System with semantic methods, *J Am Med Inform Assoc,* vol. 5, pp. 41-51, Jan-Feb 1998.

[46] K. E. Campbell, M. S. Tuttle, and K. A. Spackman, A "lexically-suggested logical closure" metric for medical terminology maturity, *Proceedings of AMIA Annual Symposium,* pp. 785-9, 1998.

[47] E. A. Mendonca, J. J. Cimino, K. E. Campbell, and K. A. Spackman, Reproducibility of interpreting "and" and "or" in terminology systems, *Proceedings of AMIA Annual Symposium,* pp. 790-4, 1998.

[48] O. Bodenreider, Circular hierarchical relationships in the UMLS: etiology, diagnosis, treatment, complications and prevention, *Proceedings of AMIA Annual Symposium,* pp. 57-61, 2001.

[49] J. J. Cimino, H. Min, and Y. Perl, Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus, *J Biomed Inform,* vol. 36, pp. 450-61, Dec 2003.

[50] J. J. Cimino, Battling Scylla and Charybdis: the search for redundancy and ambiguity in the 2001 UMLS metathesaurus, *Proceedings of AMIA Annual Symposium,* pp. 120-4, 2001.

[51] H. Gu, Y. Perl, G. Elhanan, H. Min, L. Zhang, and Y. Peng, Auditing concept categorizations in the UMLS, *Artif Intell Med,* vol. 31, pp. 29-44, May 2004.

[52] Y. Perl, Z. Chen, M. Halper, J. Geller, L. Zhang, and Y. Peng, The cohesive metaschema: a higher-level abstraction of the UMLS Semantic Network, *J Biomed Inform,* vol. 35, pp. 194-212, Jun 2002.

[53] W. Ceusters, B. Smith, A. Kumar, and C. Dhaen, Mistakes in medical ontologies: where do they come from and how can they be detected?, *Stud Health Technol Inform,* vol. 102, pp. 145-63, 2004.

[54] W. Ceusters, B. Smith, A. Kumar, and C. Dhaen, Ontology-based error detection in SNOMED-CT, *Stud Health Technol Inform,* vol. 107, pp. 482-6, 2004.

[55] O. Bodenreider, B. Smith, A. Kumar, and A. Burgun, Investigating subsumption in SNOMED CT: an exploration into large description logic-based biomedical terminologies, *Artif Intell Med,* vol. 39, pp. 183-95, Mar 2007.

[56] O. Bodenreider, B. Smith, A. Kumar, and A. Burgun, Investigating Subsumption in DL-Based Terminologies: A Case Study in SNOMED CT, presented at the First Int'l Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004), Whistler, Canada, 2004.

[57] G. Jiang and C. G. Chute, Auditing the semantic completeness of SNOMED CT using formal concept analysis, *J Am Med Inform Assoc,* vol. 16, pp. 89-102, Jan-Feb 2009.

[58] G. Wade and S. T. Rosenbloom, The impact of SNOMED CT revisions on a mapped interface terminology: terminology development and implementation issues, *J Biomed Inform,* vol. 42, pp. 490-3, Jun 2009.

[59] Y. Wang, M. Halper, D. Wei, H. Gu, Y. Perl, J. Xu, G. Elhanan, Y. Chen, K. Spackman, J. Case, and G. Hripcsak, Auditing complex concepts of SNOMED

using a refined hierarchical abstraction network, submitted to *Journal of Biomedical Informatics*, 2011.

[60] D. Wei, M. Halper, G. Elhanan, Y. Chen, Y. Perl, J. Geller, and K. A. Spackman, Auditing SNOMED relationships using a converse abstraction network, *Proceedings of AMIA Annual Symposium,* vol. 2009, pp. 685-9, 2009.

[61] D. Wei and O. Bodenreider, Using the abstraction network in complement to description logics for quality assurance in biomedical terminologies-a case study in SNOMED CT, *Medinfo*, pp.1070-1074, Cape Town, South Africa, Sep. 2010.

[62] D. Wei, Y. Wang, Y. Perl, J. Xu, M. Halper, and K. A. Spackman, Complexity measures to track the evolution of a SNOMED hierarchy, *Proceedings of AMIA Annual Symposium,* pp. 778-82, 2008.

[63] Y. Wang, M. Halper, D. Wei, Y. Perl, and J. Geller, Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED, *Journal of Biomedical Informatics,* 2010.

[64] F. Baader, *The description logic handbook : theory, implementation, and applications*. Cambridge, UK ; New York: Cambridge University Press, 2003.

[65] S. Even, *Graph Algorithms*. Potomac, MD: Computer Science Press, 1979.

[66] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press, 1993.

[67] P. I. Good, *Permutation, Parametric, and Bootstrap Tests of Hypotheses: A Practical Guide to Resampling*, 3rd ed. New York, NY: Springer, 2005.

[68] Y. Wang, D. Wei, J. Xu, G. Elhanan, Y. Perl, M. Halper, Y. Chen, K. A. Spackman, and G. Hripcsak, Auditing complex concepts in overlapping subsets of SNOMED, *Proceedings of AMIA Annual Symposium,* pp. 273-7, 2008.

[69] *College of American Pathologists - CAP Home*. [cited March 30, 2011];
Available: http://www.cap.org

[70] M. Halper, Y. Wang, H. Min, Y. Chen, G. Hripcsak, Y. Perl, and K. A. Spackman, Analysis of error concentrations in SNOMED, *Proceedings of AMIA Annual Symposium,* pp. 314-8, 2007.

[71] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL: CRC, 1993.

[72] R. H. Dolin, J. E. Mattison, S. Cohn, K. E. Campbell, A. M. Wiesenthal, B. Hochhalter, D. LaBerge, R. Barsoum, J. Shalaby, A. Abilla, R. J. Clements, C. M.

Correia, D. Esteva, J. M. Fedack, B. J. Goldberg, S. Gopalarao, E. Hafeza, P. Hendler, E. Hernandez, R. Kamangar, R. A. Kahn, G. Kurtovich, G. Lazzareschi, M. H. Lee, T. Lee, D. Levy, J. Y. Lukoff, C. Lundberg, M. P. Madden, T. L. Ngo, B. T. Nguyen, N. P. Patel, J. Resneck, D. E. Ross, K. M. Schwarz, C. C. Selhorst, A. Snyder, M. I. Umarji, M. Vilner, R. Zer-Chen, and C. Zingo, Kaiser Permanente's Convergent Medical Terminology, *Stud Health Technol Inform,* vol. 107, pp. 346-50, 2004.

[73] M. J. Lincoln, S. H. Brown, V. Nguyen, T. Cromwell, J. Carter, M. Erlbaum, and M. Tuttle, U.S. Department of Veterans Affairs Enterprise Reference Terminology strategic overview, *Medinfo,* vol. 11, pp. 391-395, 2004.

[74] H. H. Gu, G. Hripcsak, Y. Chen, C. P. Morrey, G. Elhanan, J. Cimino, J. Geller, and Y. Perl, Evaluation of a UMLS auditing process of semantic type assignments, *Proceedings of AMIA Annual Symposium,* pp. 294-8, 2007.

[75] F. Baader, D. Calvanese, D. McGuiness, D. Nardi, and P. Patel-Scheneider, Eds., *The description logic handbook : theory, implementation, and applications*. Cambridge, UK ; New York: Cambridge University Press, 2003.

[76] A. L. Rector and S. Brandt, Why do it the hard way? The case for an expressive description logic for SNOMED, *J Am Med Inform Assoc,* vol. 15, pp. 744-51, Nov-Dec 2008.

[77] D. Wei, M. Halper, G. Elhanan, Y. Chen, Y. Perl, J. Geller, and K. Spackman, Auditing SNOMED relationships using a converse Abstraction Network, *Proceedings of AMIA Annual Symposium,* pp. 685-689, 2009.

[78] K. A. Spackman, Rates of change in a large clinical terminology: three years experience with SNOMED Clinical Terms, *AMIA Annu Symp Proc,* pp. 714-8, 2005.

[79] K. A. Spackman, Normal forms for description logic expressions of clinical concepts in SNOMED RT, *Proceedings of AMIA Annual Symposium,* pp. 627-631, 2001.

[80] S. Schulz, K. Mark, and B. Suntisrivaraporn, Formal representation of complex SNOMED CT expressions, *BMC Medical Informatics and Decision Making,* vol. 8, 2008.