A Descriptive Delta for Identifying Changes in SNOMED CT

Christopher Ochs, Yehoshua Perl, Gai Elhanan

Department of Computer Science
New Jersey Institute of Technology
Newark, NJ, USA
{cro3, perl, elhanan}@njit.edu

James T. Case

National Library of Medicine (NLM)
National Institutes of Health (NIH)
Bethesda, MD, USA
james.case@mail.nih.gov

Abstract— **SNOMED CT is a large and complex medical terminology. Thousands of editing operations are applied to its content for each new release. Understanding what changed in a release is important for the end user and SNOMED CT editors. Each SNOMED CT release comes with release notes that provide a brief description of the changes that occurred and a set of "delta" files that identify individual changes in the content. The release notes are brief and changes to thousands of concepts may be described in a few sentences, whereas the delta files contain tens of thousands of individual changes. To better identify how SNOMED CT content changes between releases we introduce a methodology of creating a *descriptive delta* that captures the editing operations that were applied to SNOMED CT content in a given release in a more comprehensible form. We use this methodology to analyze editing operations that were part of a recent remodeling effort of the *Congenital disease* and *Infectious disease* subhierarchies in the large *Clinical finding* hierarchy.**

Keywords—SNOMED CT; terminology change analysis; terminology change tracking; remodeling tracking

I. Introduction

SNOMED CT [1] is a large medical terminology developed by the International Health Terminology Standards Development Organisation (IHTSDO). As of the Jan 2016 release, it contains over 319,000 active concepts and over 1.5 million relationships. In the Jan 2015, July 2015, and Jan 2016 releases of SNOMED CT over 192,000 changes affected its content. These changes included the addition of new concepts, the retiring of concepts that are no longer valid, and the modification of existing concepts (e.g., the addition and removal of attribute relationships).

From the perspective of an end user, it is important to know how (and where) SNOMED CT content is changing. Applications that utilize SNOMED CT will be affected by the changes made between each release. The IHTSDO provides release notes with each SNOMED CT release that briefly describe the major changes that occurred. For example, the Jan 2016 release notes mention that "A project to improve the consistency and quality of the Infectious disease hierarchy has begun with the remodeling of over 2,000 of the estimated 6,400 concepts that are subtypes of *Infectious disease*." However, typically, no details are provided about which concepts were modified and how they were modified.

The *Release Format 2* ("RF2") release [2] of SNOMED CT comes with a set of "delta" files that identify all atomic additions and removals that occurred from the previous release. However, the delta files are not "human readable" and they often list tens of thousands of atomic changes. Furthermore, in the delta files, one editing operation, as applied by a SNOMED CT editor, will often be represented as multiple entries in the delta file (see Fig 1b). The delta files express the minimal set of atomic additions and removals that can be applied to obtain the current SNOMED CT release from the previous SNOMED CT release. End users, however, are typically more interested in the overall changes to a concept and not the incremental, atomic ones.

In this paper we describe the creation of a human readable *descriptive delta*, based on SNOMED CT delta files, to provide a list of editing operations that were applied to individual concepts. A descriptive delta provides the semantics for the changes listed in the delta files, better fitting the needs of the initiators of the change (i.e., SNOMED CT editors) and the observers of the change (i.e., the end users of SNOMED CT). The descriptive delta methodology can be applied to the entirety of SNOMED CT, a specific subhierarchy of interest, or an individual concept. We use this methodology to investigate the editing operations that were applied to the *Clinical finding* hierarchy throughout 2015 (i.e., in the July 2015 and Jan 2016 versions of the hierarchy). We identify and characterize significant changes in the *Infectious disease* and *Congenital disease* subhierarchies that resulted from a concentrated effort to remodel these hierarchies due to errors and inconsistencies in their content.

II. Background

SNOMED CT [1] is a large, densely connected clinical terminology. SNOMED CT is used to support encoding of clinical data in Electronic Healthcare Records (EHRs), among other applications. Lee et al. [3] survey use of SNOMED CT and Elhanan et al. [4] describes a survey of SNOMED CT users. SNOMED CT's concepts are organized into a subsumption hierarchy via *Is a* relationships and are further defined using *attribute relationships* (e.g., *Associated morphology* and *Finding site*). SNOMED CT concepts are separated into 19 mostly disjoint hierarchies (e.g., *Clinical finding* and *Procedure*).
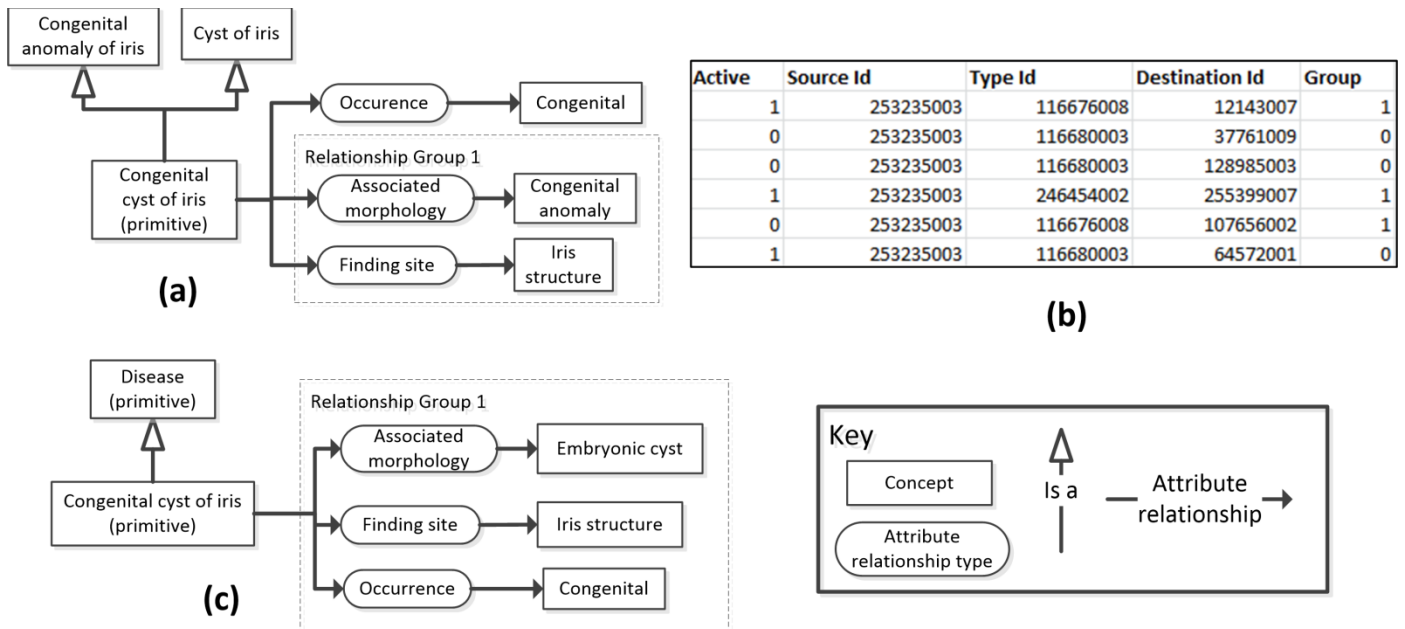
Fig. 1. (a) The stated modeling of *Congenital cyst of iris* in the Jan 2015 release. (b) The entries for *Congenital cyst of iris* in the July 2015 stated relationships delta file. (c) The stated modeling in the July 2015 release. *Note:* Relationships in Group 0 are considered *non-grouped* in RF2 format.

Fig. 1(b):

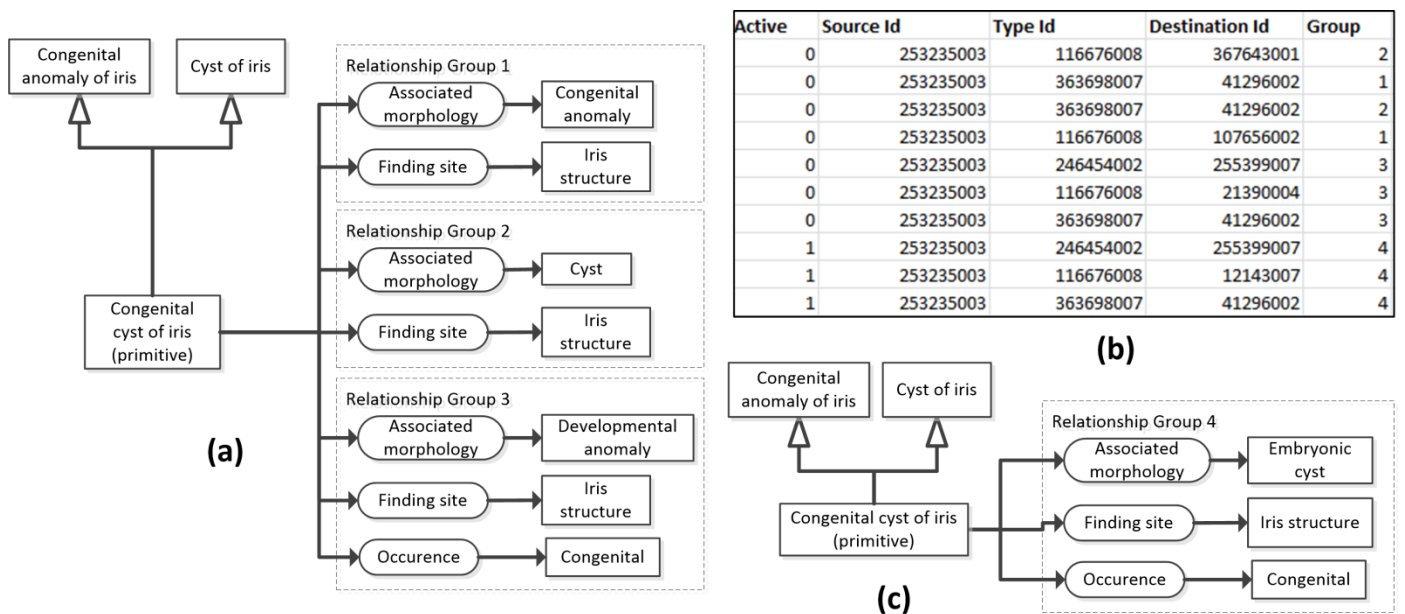| Active	Source Id	Type Id	Destination Id	Group
1	253235003	116676008	12143007	1
0	253235003	116680003	37761009	0
0	253235003	116680003	128985003	0
1	253235003	246454002	255399007	1
0	253235003	116676008	107656002	1
1	253235003	116680003	64572001	0

Fig. 2. (a) The inferred version of *Congenital cyst of iris* in the Jan 2015 release. (b) The entries for *Congenital cyst of iris* in the July 2015 inferred relationships delta file. (c) The inferred version in the July 2015 release.

Fig. 2(b):

Active	Source Id	Type Id	Destination Id	Group
0	253235003	116676008	367643001	2
0	253235003	363698007	41296002	1
0	253235003	363698007	41296002	2
0	253235003	116676008	107656002	1
0	253235003	246454002	255399007	3
0	253235003	116676008	21390004	3
0	253235003	363698007	41296002	3
1	253235003	246454002	255399007	4
1	253235003	116676008	12143007	4
1	253235003	363698007	41296002	4

SNOMED CT is released in a format called RF2 (for details and a review of this format see [2]), which consists of a set of tab-delimited text files that contain entries for concepts, relationships, and descriptions (i.e., synonyms). These files are created from the DL descriptions defined by SNOMED CT's editors. Each RF2 release comes with a "snapshot" of the current version of the terminology, the list of individual changes that occurred between the current release and the previous release (referred to as "delta" files, see examples in Fig 1b and Fig2b), and a full history of the terminology's content.

SNOMED CT is based on description logics (specifically, the EL subset of DL which allows only existential qualifiers). The *stated* version of SNOMED CT includes all of the relationships that were defined by SNOMED CT editors. The *inferred* version of the terminology is obtained by applying a classifier on the stated relationships. Each RF2 release includes the stated relationships and the relationships inferred by the classifier.

A concept may be *primitive* or *fully defined*. A concept is primitive if its defining characteristics are not sufficient to

uniquely distinguish its meaning from other similar concepts. A concept is fully defined if its defining characteristics are sufficient to distinguish its meaning from other similar concepts. If a concept is fully defined, the SNOMED CT classifier can add inferred hierarchical relationships according to the logical definitions of the concepts.

A concept may be modeled using several attribute relationships of the same type. For example, in Fig 2a *Congenital cyst of iris* has three *Associated morphology* and *Finding site* attribute relationships. To clarify the meaning of a group of attribute relationships, SNOMED CT provides a mechanism of *relationship groups* [5] to clarify the definition of a concept. Specifically, relationship groups combine multiple attribute relationships into a single unit.

Ceusters et al. [6] reviewed the changes across 18 releases of SNOMED CT. They found that SNOMED CT history mechanism did not sufficiently provide justifications for stated changes. Ceusters [7] also tracked SNOMED CT evolution to determine its quality. In previous studies, Rector et al. [8], Mortensen et al. [9], and Ochs et al. [10] have investigated quality assurance issues in the *Clinical finding* hierarchy's content. Other significant quality issues in *Clinical finding* have also been reported, leading to the current project to remodel the *Congenital disease* and *Infectious disease* subhierarchies.

III. METHODS

A. Congenital Disease and Infectious Disease Remodeling

As noted in SNOMED CT July 2015 and Jan 2016 release notes, the *Congenital disease* and *Infectious disease* subhierarchies are being remodeled to address inconsistencies and errors in their content. Both subhierarchies each contain over 6,000 concepts and are in the large *Clinical finding* hierarchy, which has 103,912 concepts in the Jan 2016 release. While this editing effort was initiated by the IHTSDO, change requests may be submitted by individual users or SNOMED CT national release centers.

Many SNOMED CT concepts follow the *closest proximal primitive parent* stated concept model [11]. A proximal primitive parent is the closest parent in the hierarchy to a concept being modeled that *is not* fully defined. The goal of the remodeling effort is to correct inconsistent stated modeling to enable the classifier to infer most of the hierarchical connections between the concepts in these subhierarchies. The remodeling effort is still ongoing and it is being carried out by co-author (JTC), IHTSDO Head of Terminology (acting) and the content manager of the US Extension of SNOMED CT.

For example, in Fig 1a, *Congenital cyst of iris* has two stated parents, *Congenital anomaly of iris* and *Cyst of iris*, both fully defined concepts. Thus, this concept does not follow the proximal primitive parent. In Fig 1c, which shows the stated modeling of *Congenital cyst of iris* in the July 2015 release, the concept has just one stated parent, *Disease*, the proximal primitive parent as of the Jan 2016 release.

A second issue is inconsistent grouping of attribute relationships. In the proximal primitive parent model, attribute relationships must be properly assigned to relationship groups. For example, in Fig 1a *Congenital cyst of iris*'s *Occurrence* attribute relationship was not assigned to a relationship group. In the new stated modeling of the attribute relationships are all assigned to the same relationship group.

As of the Jan 2015 release, the *Congenital disease* hierarchy had 4,566 concepts with a stated fully defined parent and 2,466 concepts with a stated attribute relationship that wasn't in a relationship group. Similarly, in the *Infectious disease* hierarchy 1,065 concepts had a stated fully defined parent and 4,198 had a stated attribute relationship in no relationship group.

By modeling concepts using the closest proximal primitive parent design pattern, the SNOMED CT classifier can more accurately infer parents and descendants and remove the incorrect inferences. Furthermore, SNOMED CT editors will only need to assign the proper proximal primitive parent. There will be no need to try and determine where the concept should be in the inferred hierarchy. The other advantage of this editing approach is that it has allowed a substantial number of primitive concepts to become fully defined.

B. Descriptive Delta

We will now describe a methodology to extract a human readable descriptive list of editing operations that were applied between two SNOMED CT releases. We call this list a "descriptive delta." This process utilizes the delta files provided with each SNOMED CT release. Our focus is on capturing the editing operations that were applied to the relationships used in concept definitions.

A SNOMED CT relationship can be represented as a (*source concept*, *relationship type*, *destination concept*, *relationship group*) tuple. For a relationship r_i this can be abbreviated as (s_i, t_i, d_i, g_i). The relationship delta files (for stated relationships and inferred relationships) also indicate if the given relationship is *active* (i.e., added) or inactive (i.e., removed) in the current release (see Fig 1b, Fig 2b).

Given the set of active and inactive relationships, as expressed in a relationship delta file, and the inferred concept hierarchy for the current release, the editing operations that were applied to the active concepts can be derived. Below we identify five important examples of editing operations and how they are identified using the delta file.

Added relationship: Given an activated relationship r_1 there exists no retired relationship r_2 where $s_1 = s_2$, $t_1 = t_2$, and $d_1 = d_2$. d_1 is also not an ancestor or descendant of d_2 in the current release's inferred concept hierarchy.

Removed relationship: Given a retired relationship r_1 there exists no active relationship r_2 where $s_1 = s_2$, $t_1 = t_2$, and $d_1 = d_2$. d_1 is also not an ancestor or descendant of d_2 in the current release inferred concept hierarchy.

Relationship Target More Refined: Given an inactive relationship r_1 there exists an active relationship r_2 where

$s_1 = s_2$, $t_1 = t_2$, and d_2 is a descendant of d_1 in the current release's inferred concept hierarchy.

Relationship Target Less Refined: Given an inactive relationship r_1 there exists an activated relationship r_2 where $s_1 = s_2$, $t_1 = t_2$, and d_2 is an ancestor of d_1 in the current release's inferred concept hierarchy.

Relationship Group Changed: Given an active relationship r_1 there exists an inactive relationship r_2 where $s_1 = s_2$, $t_1 = t_2$, and $d_1 = d_2$ but g_1 does not equal to g_2.

The process of converting a delta file into a descriptive delta may be a one-to-one process, where one delta entry is captured as one editing operation (e.g., for an added relationships), or a many-to-one process (e.g., for relationship target more refined).

The first four editing operations can be further refined by distinguishing between changes to *Is a* relationships and changes to attribute relationships (e.g., ***Added parent*** versus ***Added attribute relationship*** and ***More refined parent*** versus ***Attribute relationship target more refined***). This distinction provides a separation between changes in the hierarchy and changes in the relationships used to define each concept.

We note that it is possible to capture additional kinds of editing operations, and further refinements of the above editing operations, when necessary. If a concept has only one parent, and there exists both an inactive *Is a* relationship for that concept and an active *Is a* relationship for that concept, and the new parent is not an ancestor or descendant of the old parent in the inferred hierarchy, this could be identified as a ***changed parent*** editing operation. For example, the concept *Midline sinus of the upper lip* had one parent, *Lip pits*, in the Jan 2015 release. This concept was deactivated in July 2015 and *Midline sinus of the upper lip* parent was changed to *Congenital lip pits*. However, in this study we focus on the core editing operations listed above.

Two or more editing operations may relate to the same modeling change. For example, in Fig 1, the two stated parents of *Congenital cyst of iris*, *Congenital anomaly of iris* and *Cyst of iris*, were replaced by a single stated parent, *Disease* (the proximal primitive parent), an ancestor of both of the inactivated parents (only the *Is a* relationships were inactivated in this example, not the parent concepts themselves). Furthermore, a given concept may have multiple instances of a given kind of editing operation. For example, in its stated modeling, *Congenital cyst of iris* has two instances of stated attribute relationships being added.

Applying this process on a stated relationships delta file identifies the changes that were intentionally applied by a SNOMED CT editor. Consider the changes listed in the stated relationships delta file for *Congenital cyst of iris*, as illustrated in Fig 1c. In the descriptive delta, these are identified as:

Less refined parent (2): *Congenital anomaly of iris* and *Cyst of iris* were replaced by *Disease.*

Added attribute relationships (2): *Associated morphology* with a target of *Embryonic cyst* and *Occurrence* with a target of *Congenital*.

Removed attribute relationship (1): *Associated morphology* with a target of *Congenital anomaly*

Similarly, the descriptive delta process can be applied on the inferred relationships delta file. This will highlight the implicit changes that occurred between the inferred versions of the concepts. Consider the changes listed in the July 2015 inferred relationships delta file for *Congenital cyst of iris*, as shown in Fig 1d. These implicit changes are captured as:

More refined attribute relationship target (2): *Associated morphology* with a target of *Developmental anomaly* and *Associated morphology* with a target of *Cyst* replaced by *Associated morphology* with a target of *Embryonic cyst*.

Removed attribute relationship (3): *Associated morphology* with target *Congenital anomaly* and two *Finding site* relationships with target *Iris structure*.

Relationship group changed: Finding site with target *Iris structure* moved to relationship group 4.

The descriptive delta for a stated relationships file and the descriptive delta for an inferred relationships file will often be very different, as the classifier will make significant changes based on the stated relationships. A concept that had no stated editing operations applied may have several implicit changes due to editing operations at its ancestor(s). Given a concept, one can use the inferred concept hierarchy to identify which ancestors were modified and which editing operations were applied at those ancestors. Alternatively, a concept may have stated editing operations applied but this does not affect the concepts in the inferred version of SNOMED CT, and thus, it would have no implicit changes in the descriptive delta created from the inferred delta file.

There may be tens of thousands of editing operations identified. To reduce the amount of information that has to be reviewed, a user can focus on only a subset of the content that was edited. For example, a user can choose a concept and look at only the editing operations that affected the subhierarchy rooted at that chosen concept. Alternatively, they can obtain a summary of the descriptive delta by only looking at the number of editing operations applied. For example, the stated descriptive delta for *Congenital cyst of iris* could be summarized as "Less refined parent (2), Added attribute relationship (2), Removed attribute relationship (1)."

If a user is interested in identifying which portions of a SNOMED CT hierarchy underwent significant change they could review the number of concepts that were affected by stated editing operations in certain subhierarchies (e.g., a user can see that relatively many editing operations occurred in the *Congenital disease* and *Infectious disease* subhierarchies in July 2015, see Results).

To create descriptive deltas we have developed a software tool that takes as input a relationship delta file and a concept

of interest *c*. The output identifies all of the editing operations that affected the subhierarchy rooted at *c*.

In the initial version of this tool the output is presented tab-delimited text, where one line consists of the concept identifier, the name of the concept, the kind of editing operation applied, and a textual description of the change (e.g., "Parent *Cyst of iris* was replaced by *Disease*"). This output can be inserted into an Excel spreadsheet to obtain aggregated information. In Discussion we mention a future version of this tool which will provide a visual change browser.

IV. RESULTS

We obtained the July 2015 and Jan 2016 International SNOMED CT releases in RF2 format and created descriptive deltas using the stated relationships delta file and inferred relationships delta file provided in each release. We focused our analysis on concepts that changed, as opposed to concepts that were added and removed, since we concentrated our attention to the remodeling efforts.

First, we investigated the editing operations that were applied to the concepts in the complete *Clinical finding* hierarchy. The July 2015 and Jan 2016 stated relationships delta files contained 7,737 and 10,190 entries for the *Clinical*

finding hierarchy, respectively. The inferred relationships delta files contained 33,146 and 39,380 entries, respectively. We identified 5,728 and 7,524 descriptive delta editing operations in the July 2015 and Jan 2016 release, respectively. We identify which subhierarchies (rooted at the children of *Clinical finding* and *Disease*) had the most edited concepts.

Table 1. The five *Clinical finding* and *Disease* subhierarchies containing the most concepts that had at least one stated editing operation applied, along with their status in Jan 2016. The percentage of the concepts in the subhierarchy that was edited is shown in parenthesis.

	# Concepts Edited in July 2015 (%)	# Concepts Edited in Jan 2016 (%)
Clinical finding subhierarchies		
Disease	2,586 (3.76%)	2,826 (4.08%)
Finding by site	2,435 (3.87%)	2,004 (3.15%)
Clinical history and observation findings	108 (0.6%)	363 (2.05%)
Bleeding	55 (5.06%)	37 (3.25%)
Deformity	48 (5.93%)	8 (1.00%)
Disease subhierarchies		
Disorder by body site	2,224 (4.37%)	1,924 (3.76%)
Infectious disease	893 (14.4%)	1,815 (29.1%)
Inflammatory disorder	759 (10.4%)	855 (11.5%)
Congenital disease	672 (10.2%)	20 (0.3%)
Traumatic AND/OR non-traumatic injury	151 (1.58%)	235 (2.46%)

Table 2. Metrics from the descriptive deltas created for the *Clinical finding* hierarchy and its *Congenital disease* and *Infectious disease* subhierarchies.

	Clinical finding		Congenital disease		Infectious disease	
	# Concepts in July 2015	*# Concepts in Jan 2016*	*# Concepts in July 2015*	*# Concepts in Jan 2016*	*# Concepts in July 2015*	*# Concepts in Jan 2016*
Active	103,227	103,912	6,616	6,614	6,221	6,240
Primitive	58,799	58,572	4,276	4,266	2,104	1,770
Added	1,176	962	52	7	169	128
Removed	167	277	36	9	47	109
Stated Relationships Descriptive Delta Metrics (# Concepts with the given operation applied)						
Edited (any operation)	1,738	2,395	620	13	727	1,687
Added parent	184	122	36	3	27	19
Removed parent	330	229	98	3	141	120
More refined parent	48	43	15	2	8	9
Less refined parent	1,153	1,861	405	4	587	1,488
Added attribute relationship	1,383	1,843	529	9	705	1,592
Removed attribute relationship	243	288	162	4	17	46
More refined attribute relationship target	120	150	52	0	31	110
Less refined attribute relationship target	180	50	143	0	20	30
Relationship group changed	305	401	136	0	173	389
Inferred Relationships Descriptive Delta Metrics (# Concepts with the given operation applied)						
Changed (any operation)	7,132	13,065	1,098	101	4,261	4,395
Added parent	929	1,114	219	28	251	509
Removed parent	1,264	1,167	126	23	862	646
More refined parent	752	793	192	14	81	368
Less refined parent	823	611	92	3	664	406
Added attribute relationship	3,176	8,502	271	41	2,531	1,475
Removed attribute relationship	1,289	1,941	601	18	80	1,374
More refined attribute relationship target	475	1,123	329	11	42	827
Less refined attribute relationship target	446	137	149	4	257	81
Relationship group changed	4,481	4,114	836	23	3,327	3,830

Table 3. Three examples of edited concepts from the stated descriptive delta.

Concept	Stated editing operations
Congenital tuberculosis	Added attribute relationships *Causative agent* with target *Mycobacterium tuberculosis complex*, *Pathological process* with target *Infectious process*, and *Occurrence* with target *Congenital*

Urosepsis	Removed parent
	   Urinary tract infectious disease
	
	Added attribute relationship
	   Due to with target *Urinary tract infectious disease*

Zika virus disease	Less refined parent
	   Viral disease replaced with *Disease*
	   (proximal primitive parent)
	
	Added attribute relationships
	   Causative agent with target *Zika virus* and *Pathological process* with target *Infectious process*

In both releases the majority of edited concepts were in the *Disease* and *Finding by site* subhierarchies (Table 1). This is not surprising, as these subhierarchies both contain over 60,000 concepts and both contain many of the same concepts (i.e., there are many diseases that can be found in many body sites). The remaining subhierarchies had few changes. The effect of the remodeling of *Congenital disease* and *Infectious disease* is reflected in Table 1.

About 1,500 of the 2,586 edited *Disease* concepts (over 58%) were in the *Congenital disease* and *Infectious disease* subhierarchies (these are mostly disjoint; they share only 75 concepts). On the other hand, the *Disorder by body site* and *Inflammatory disease* subhierarchies contain relatively many concepts that are also in either the *Infectious disease* or *Congenital disease* subhierarchies.

Table 2 provides metrics from the descriptive deltas for July 2015 and Jan 2016. It lists the number of concepts that had a particular type of editing operation applied. We found that an edited concept typically undergoes multiple editing operations. Each edited concept had, on average, 3-4 editing operations applied. The most common editing operations applied in these hierarchies was making a parent more general (i.e., assigning the proximal primitive parent), adding attribute relationships, and assigning relationship groups to existing stated attribute relationships. These editing operations are in-line with the goals of the remodeling effort being undertaken in these hierarchies. Table 3 provides three examples of the stated modeling of concept changing.

V. DISCUSSION

Identifying changes between two versions of SNOMED CT is important for end users and for SNOMED CT editors. The descriptive delta described in this paper provides a human readable, actionable format that is not currently available. As analogy, the delta files provided in SNOMED CT release are like machine code and the descriptive delta is like a higher level programming language.

Sometimes a user realizes certain changes occurred in a SNOMED CT subhierarchy that they are interested in but they do not know the nature and motivation for the changes. The descriptive delta, with its semantic view of change, can better support such a user in finding out the nature of the changes.

The descriptive delta analysis reflects the extensive work that went into remodeling portions of the *Clinical finding* hierarchy in July 2015 and Jan 2016. Namely, thousands of editing operations were applied manually to the concepts in the *Congenital disease* and *Infectious disease* subhierarchies. These changes affected thousands of concepts directly and thousands more implicitly (Table 2). Whenever this much change occurs one needs to assess the overall impact (and correctness) of the changes. During this study we identify concepts that underwent significant changes (e.g., *Congenital heart block*) but still were not yet completely remodeled.

From Tables 1 and 2 we notice that concerted editing efforts have a detectable impact on SNOMED CT's content. For example, the *Congenital disease* subhierarchy underwent extensive remodeling in July 2015 but almost no remodeling in Jan 2016. In comparison, the *Infectious disease* subhierarchy saw significant remodeling in both releases. The *Congenital disease* remodeling project is still in progress, but the focus of the Jan 2016 editing cycle was on *Infectious diseases*. One limitation is the amount of resources (in terms of editor's time) that can be dedicated to these large remodeling projects.

From Table 2 two we observe that the number of primitive concepts in *Infectious disease* has decreased significantly. This is the desired result of the remodeling effort. In the *Congenital disease* subhierarchy, however, there was no significant reduction. Many concepts that were remodeled could not be fully defined due to limitations in the concept model (e.g., lack of *finding site* and *associated morphology* target values of sufficient specificity, SNOMED CT does not currently contain concepts specific enough to be proper targets). From Table 2 we also see that thousands of concepts were affected by inferred changes. By removing the stated fully defined parents from concepts in these subhierarchies, and remodeling the concepts using the proximal primitive parent, the classifier significantly affected the contents.

In future work we will continue to observe the effects of the remodel effort being applied to the *Congenital disease* and *Infectious disease* subhierarchies. We will perform a review of sets of concepts that underwent remodeling to evaluate the correctness of the changes. We will compare the results of this review with the output of the descriptive delta to determine if concepts that undergo certain editing operations are more prone to needing additional remodeling.

While our current descriptive deltas are output in a tab-delimited file format, we plan to provide them in an interactive application-based display that will enable end users to easily investigate and search the changes based on the type of change, areas of interest, specific concepts, and changes in the inferred version resulting from stated changes. We will also investigate combining the descriptive delta technique with diff partial-area taxonomies [12], which we developed to visually summarize change between two releases of an

ontology. Using this method we could create a "heat-map" of where similar SNOMED CT content is changing.

VI. Conclusions

In this paper we introduced a method of creating descriptive deltas for SNOMED CT to identify the editing operations that were applied to its concepts. The descriptive delta helps to identify the nature of the remodeling the large amount of change in the *Clinical finding* hierarchy that occurred due to the still-in-progress remodeling of the *Congenital disease* and I*nfectious disease* subhierarchies.

References

[1] Stearns, M. Q., Price, C., Spackman, K. A., et al. SNOMED clinical terms: overview of the development process and project status. Proc AMIA Annu Symp. 2001:662-6.

[2] Ceusters, W. SNOMED CT's RF2: Is the Future Bright? Studies in health technology and informatics. 2011;169:829-33.

[3] Lee, D., Keizer, N. d., Lau, F., et al. Literature review of SNOMED CT use. J Am Med Inform Assoc. 2014;21(e1):e11-e9.

[4] Elhanan, G., Perl, Y., Geller, J. A survey of SNOMED CT direct users, 2010: impressions and preferences regarding content and quality. J Am Med Inform Assoc. 2011;18 Suppl 1:i36-44.

[5] Cornet, R., Schulz, S. Relationship groups in SNOMED CT. Studies in health technology and informatics. 2009;150:223-7.

[6] Ceusters, W. M., Spackman, K. A., Smith, B. Would SNOMED CT benefit from realism-based ontology evolution? AMIA Annu Symp Proc. 2007:105-9.

[7] Ceusters, W. Applying Evolutionary Terminology Auditing to SNOMED CT. AMIA Annu Symp Proc. 2010;2010:96-100.

[8] Rector, A. L., Brandt, S., Schneider, T. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. J Am Med Inform Assoc. 2011;18(4):432-40.

[9] Mortensen, J. M., Minty, E. P., Januszyk, M., et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. J Am Med Inform Assoc. 2015;22(3):640-8.

[10] Ochs, C., Geller, J., Perl, Y., et al. Scalable Quality Assurance for Large SNOMED CT Hierarchies Using Subject-based Subtaxonomies. J Am Med Inform Assoc. 2014;22(3):507-18.

[11] Hernandez, P., Richardson, C. Authoring SNOMED CT: Generic Authoring Principles SNOMED CT Implementation Showcase: IHTSDO; 2014 [26 February 2016]. Available from: http://ihtsdo.org/fileadmin/user_upload/doc/showcase/show14/SnomedCtShowcase2014_Present_14107.pdf.

[12] Ochs, C., Perl, Y., Geller, J., et al. Summarizing and Visualizing Structural Changes during the Evolution of Biomedical Ontologies Using a Diff Abstraction Network. Journal Of Biomedical Informatics. 2015;56:127-44.