Using the SNOMED CT Error Taxonomy to Maximize the SNOMED CT Content Request System

NLM Associate Fellow: J. Caitlin Sticco

Project Sponsors: Dr. Olivier Bodenreider and Dr. James Case

August 2011

Table of Contents

Abstract 3

Background 3

Methods 4

Results and Discussion: SNOMED CT Error Taxonomy 5

- SNOMED CT Quality Review 5
- Separation of Errors and Causes 7

Results and Discussion: NJIT Error Analysis 10

- Highlights from Errors and Causes 12
  - Missing Attributes and Incompletion 12
  - Problematic Conceptualizations 12
  - No Apparent Cause 13
  - Suggestion Too General 14
  - Ancestral Errors 14
  - Parent Too Specific: 15
  - Parent Too General: 15
  - Parent Unrelated: 16

Recommendations 17

- Confirm findings with additional data 17
- Reduce Conceptual Ambiguity 17
  - Track concepts that engender confusion 17
  - Provide textual definitions for concepts 18
- Revise the SNOMED CT Request Submission System 18
  - Improve Usability 18
  - Include and Organize Essential System Tasks 18
  - Prompt the Submitter 20

Appendices 22

- Appendix A 22
  - SNOMED CT Error Taxonomy 22
- Appendix B 26
  - Usability Feedback on the SNOMED CT Beta Request System 26

Abstract

Background

SNOMED CT (Systematized Nomenclature of Medicine--Clinical Terms) is a comprehensive clinical terminology, currently distributed by the International Health Terminology Standards Development Organisation (IHTSDO). The National Library of Medicine (NLM) acts as the National Release Center for SNOMED CT for the US. Recently, NLM developed a new system for submitting requests for changes to SNOMED CT, the US SNOMED CT Content Request System (USCRS).

Objective

To create a taxonomy that describes the types and causes of errors occurring in SNOMED CT in order to inform the further development of the request system, as well as general research on quality assurance in SNOMED CT.

Methods

To achieve these ends, we examined requests submitted via the original email method. We also tested the SNOMED CT request submission system beta version to gain insight into its current functions and design, and we reviewed relevant literature and SNOMED CT documentation. The SNOMED CT error taxonomy was created through a gradual process of successive reviews of submissions and literature alternating with refinements to the growing error taxonomy. Eventually the taxonomy was used to formally code the requests examined.

Results

The taxonomy distinguishes clearly between observable Errors, presumed Causes for them, and the basic Actions that can remedy the Errors. It is demonstrated to be useful to identifying and describing previously submitted errors.

Conclusions

Based on the error taxonomy, changes that could be made to make the USCRS more user-friendly and comprehensive are suggested. Use of the error taxonomy could also improve the quality of requests and error reports.

Background

SNOMED CT (Systematized Nomenclature of Medicine--Clinical Terms) is a comprehensive clinical terminology, originally created by the College of American Pathologists (CAP) and currently distributed by the International Health Terminology Standards Development Organisation (IHTSDO), of which the United States is a member. The National Library of

Medicine (NLM) acts as the National Release Center for SNOMED CT for the US as well as the creation, maintenance and distribution of the US Extension to SNOMED CT and a diverse set of reference sets derived from both the International Release and the US extension. Finally, the NLM acts as the primary gateway for requests to add or change content to both the International Release and US Extension to SNOMED CT. Thus, NLM is responsible for the distribution of SNOMED CT in the US and for submitting requests from the US to the IHTSDO.

Until recently, users of SNOMED CT submitted requests via direct email to Medlars Management Section of the NLM. This system was inefficient, lacking any standardization for the content or format of the request submitted. In most cases, multiple back and forth communications were required to clarify requests before corrective action could be taken. To improve the quality of requests and the ease of their submission and review, a new system for submitting request has been created, the US SNOMED CT Content Request System (USCRS). This system was in beta phase during the research for this report, and will be in its first official release as of September 2011.

As expected with an initial release, the content request system currently has limited capabilities. The intent of this report is elucidate the types and causes of errors that are reported in SNOMED CT in order to inform the further development of the request system, as well as general research on quality assurance in SNOMED CT. In the service of this, we have created an taxonomy of SNOMED CT errors. We hope to improve the quality of new request submissions, as well as generate additional information from the submissions regarding possible additional errors or deficiencies elsewhere in SNOMED CT.

Methods

To achieve these ends, we examined 108 requests submitted via the original email method from the New Jersey Institute of Technology (NJIT) and 83 from the RadLex radiology ontology, which is currently being mapped to SNOMED CT. All requests had been previously reviewed and annotated by the NLM Health Program Specialist for SNOMED CT, Dr. James Case, with the normal notes and feedback used while processing requests. We also tested the SNOMED CT request submission system beta version to gain insight into its current functions and design, and we reviewed relevant literature and SNOMED CT documentation.

The SNOMED CT error taxonomy was created through a gradual process of successive reviews of submissions and refinements to the growing error taxonomy. Not all errors and causes were observed in our examined datasets. Instead, some were implied by others (correlated categories), and some were suggested by literature instead of the dataset.

Eventually the completed taxonomy was used to formally code each submission from the NJIT for the types of errors it contained. The RadLex submissions were not similarly coded, because they were found to be qualitatively less revealing and significantly more uniform, since they are

primarily requests for new concepts. The results reported below focus on the analysis of the dataset from the NJIT. However, the comments from RadLex review are also available in the related Access Database.

Results and Discussion: SNOMED CT Error Taxonomy

The error taxonomy we developed consists of three main classes of concepts: Errors, Actions, and Causes. Errors are problems with the SNOMED CT ontology. Actions are ways the ontology may be modified; thus, Errors can be corrected by Actions. Causes obviously cause Errors. Note that what might be casually referred to as "user errors" fall under Causes in this system. Errors refer only to observable problems within the SNOMED CT ontology itself.

All errors are intended to refer to a problem with an individual concept, such as a concept missing a parent or having an incorrect attribute. All errors in parent-child relations (i.e., parent-to-child and child-to-parent) are normalized to errors in child-to-parent relations. If a concept has an incorrect child, for example, that problem is represented as an incorrect parent of the child concept. The SNOMED CT error taxonomy is available in Appendix A.

SNOMED CT Quality Review

IHTSDO's own quality review methodology aligns closely with our proposed taxonomy. Table 3 shows the revised list of review questions from the SNOMED CT Quality Review Phase 2 Draft Methodology Criteria, Tooling and Usage [1], with the corresponding Errors from our taxonomy that they seek to find. This exposes a few review aspects missing from our error taxonomy. These are highlighted in pink in Table 1 below.

Specifically, we do not address two questions relating to synonym ambiguity: "Will the synonyms give rise to ambiguity?" and "Are any of the synonyms applied to any other active SNOMED CT concepts?" These two questions both address issues we do not believe are errors. Ambiguity in synonyms and other descriptions is an unavoidable consequence of natural language. SNOMED CT addresses this problem not by artificially limiting the number of meanings a given word might have, but by assigning different identifiers to each meaning of the word. Thus each description has single descriptionIDs assigned to a single conceptID. If the same string is assigned as a description for another concept, it will have a different description ID. Thus ambiguity is eliminated.

Table 1: SNOMED Review Questions

SNOMED CT Quality Review Phase 2 Draft Methodology Review Questions	**SNOMED CT Error Taxonomy**
I. Reviewing the descriptions for a concept	--
i. Synonyms inappropriate for clinical use	Non-clinical synonym
ii. Any missing synonyms that are clinically useful	Description missing
iii. Will the synonyms give rise to ambiguity (e.g. the same synonym belonging to different concepts)?	?

iv. Are any of the synonyms applied to any other active SNOMED CT concepts	?
v. Are there any synonyms that are wrong i.e. do not represent the same concept?	Description incorrect
vi. Is the preferred term inappropriate or could one of the synonyms be a better preferred term.	Term non-preferred
vii. Reviewing the hierarchical placement for a concept	--
i. Is the hierarchy unintuitive or incorrect?	Unrelated parent?
ii. Are there any incorrect parents?	Parent incorrect
iii. Are there any missing parents?	Parent missing
iv. Are there any incorrect children from a clinical perspective?	(Parent incorrect)
v. Are there any correct children from a clinical perspective that could be better placed in the hierarchy?	(Parent too general, too specific, or unrelated)
vi. Are there any missing children of clinical relevance?	Term missing from vocabulary or (Parent Missing)
vii. Are there any incorrect siblings from a clinical perspective?	(Parent incorect)
viii. Are there any correct siblings from a clinical perspective that could be better placed in the hierarchy?	(Parent too general, too specific, or unrelated)
ix. Are there any missing siblings of clinical relevance?	Term missing from vocabulary or (Parent Missing)
viii. Are there any other issues with this or related concepts?	--
II. Clinical modelling review (questions for clinical terminologists)	--
a. Is the Fully Specified Name incomplete or ambiguous	Description incorrect or description incomplete
b. If no, is there a more appropriate description which should be used?	Description missing
c. For the fully-defined concepts, do the defining relationships, correctly, accurately and sufficiently capture the meaning of the concept?	Any errors in parents, children, or attributes
d. For the primitive concepts, can they be changed into fully-defined concepts using approved attributes and values?	Attribute Missing or Term missing from vocabulary
e. For the primitive concepts, can they be changed into fully-defined with changes to the model or content?	Term missing from vocabulary
f. For all concepts, do they conform to the style guide/concept model?	--
i. Were you able to identify appropriate sections of the style guide/concept model/editorial guidance	--
ii. Was the style guide detailed enough to fully model the concept (may need refining?)	Any type of error caused by Inadequate concept model
iii. Are there any attributes (defining and non-defining) which could be applied and have not been applied?	Attribute missing
iv. Any attributes (defining and non-defining) that have been applied and should not have been applied?	Attribute incorrect
v. Any attributes (defining and non-defining) that have been applied inappropriately?	Attribute incorrect
III. Stylistic editorial checks	--
a. This will review the representation of description itself, based on published rules, to include the following:	--
i. Any errors in spacing	Spacing Incorrect

ii. Any errors in capitalizations	Capitalization Incorrect
iii. Any errors in plurals	Pluralization Incorrect
iv. Any errors in spelling	Spelling Incorrect
v. Any errors in abbreviations and the use of acronyms	Abbreviation Incorrect

Separation of Errors and Causes

Much of the current literature conflates errors with their causes. This makes it difficult to translate directly between the concepts our proposed error taxonomy and those in the literature. For example, Rector et al. give the following typology of errors in [2], and a similar typology in [3]:

Simple mistakes
Misunderstandings of the semantics of concepts and attributes as implemented in the description logic
Over-literal definitions
Incomplete modelling
Attempts to fix erroneous inferences without tracing them to their roots
Lack of normalisation of complex segments
Fundamental errors in the modelling schemas

However, many of these combine the types of causes and the types of errors. For example, "simple mistakes" are perhaps not a type of error in the ontology, but rather an error on the part of the user; that is, the *cause* of errors in the ontology. This is reflected even within this typology: "Simple mistakes" could lead to "incomplete modeling," for example. This kind of conflation is a problem for the practical application of these kinds of typologies, since they contain many overlapping categories. Additionally, while errors themselves are observable, causes of error will always be inferred and speculative. While performing analysis of problems, grouping problems by the actual error may be of more interest sometimes, while in other cases, grouping by cause may be more useful. For clarity and precision, it may be useful to separate the description of errors and the causes of errors, which is what we have done in our proposed taxonomy. Below, I have attempted to harmonize Rector's typology with our taxonomy.

Reconciling the Rector typology to our taxonomy exposes places where it is useful to group types of errors, or where an underlying cause tends to create errors in tandem. Because the Errors are observable, but the Causes are inferred, understanding which Errors are symptoms of which Causes could be helpful. When analyzing SNOMED CT errors, we would be better able to identify the most appropriate way to correct problems as well as what other problems may have also occurred from the same Cause.

Misunderstandings of the semantics of concepts and attributes as implemented in the description logic

Rector explains the incorrect inferences that result from systematic diseases being sited in localized tissues or organs, such as *diabetes* being given *finding_site pancreas*. He posits that

this mistake is related not to a misunderstanding of the concepts themselves, but instead to misunderstanding the semantics of *finding-site.* Editors do not always understand the semantics of how relationships are implemented in the ontology and the inferences that will result from classification. In our taxonomy, this lack of understanding is a Cause of errors in the ontology. Rector's language "semantics of concepts" suggests a misunderstanding of concept definitions, which is not the case, so we have instead called this *Misuse of semantics of description logic*.

Over-literal definitions

Over-literal definitions leading to over-generalized concepts is a combination of error cause and type. One example given is *subdural hemorrhage*. This term is used clinically to *imply intracranial subdural hemorrhage*, unless it is specified as *spinal subdural hemorrhage*. However, in SNOMED CT it had been logically defined to mean either. Furthermore, *intracranial subdural hemorrhage* had not been given as a concept at all, and *subdural hemorrhage* was simply a parent to *spinal subdural hemorrhage*.

According to our taxonomy, this would be a combination of errors. If *subdural hemorrhage* actually means *intracranial subdural hemorrhage*, then it is *Parent too specific* for *spinal subdural hemorrhage*. Additionally, a concept is missing that could represent both types of subdural hemorrhage, so Concept *missing* would also be an error here. These errors have both arisen from the problematic conceptualization of *subdural hemorrhage*, it appears from an over-literal definition. We have added over-literal definition as one of the subtypes of incorrect conceptualization.

Incomplete modeling

The description logic used for SNOMED CT, EL+, editors can choose to delineate a class as defined (with necessary and sufficient conditions) or primitive (with necessary conditions only). However, the classifier can only classify defined classes. While there are legitimate reasons not to fully define classes, there are also some classes that should be defined that have instead been as left as primitives. This is an error, and it propagates others. Due to the presence of inappropriate primitive concepts, not all appropriate inferences are made. Unlike most other errors in our taxonomy, this is not an error in the relationships between the concepts (although it may give rise to them elsewhere), but rather an error purely in the application of description logic. We have called this *Missing full definition despite the presence of necessary conditions.*

Attempts to fix erroneous inferences without tracing them to their roots

Rector et al. describe users attempting to fix missing inferences by asserting them manually wherever they are noticed missing. Since many missing inferences can stem from a single missing relationship higher in the hierarchy, these fixes are generally patchy, what Rector calls "helter-skelter modeling." Additionally, once the originating, top-level missing relationship is created, the manually asserted axioms lower in the hierarchy will become redundant. We considered adding these redundancies to the taxonomy as *Parent redundant* and *Attribute redundant* errors. However, we concluded that redundancies in assertion and inference are not

errors in themselves. Rector argues that these redundancies should be removed to ensure that future changes propagate correctly through the hierarchy, without leaving behind old assertions that are now incorrect. While this is true, and we consider removal of these redundancies a maintenance procedure, not a corrective one—prevention of future errors, not removal of current errors. We have however, added the process of redundancies from local fixes becoming errors as a Cause: local over-assertion.

Lack of normalisation of complex segments

This describes the underlying cause of a group of errors. When many identical facets are used to describe related terms in a branch, (for example the site, severity, stage, and symptom of a disease process), these facets may be applied haphazardly to concepts as the branch develops over time. The lack of consistency in the use of such definitional axes "weakens" the ontology by making it more prone to errors and omissions, as well as to inconsistency in definitions of terms that share facets. This Cause was difficult to incorporate into our taxonomy, as it represents a systematic problem, as opposed to the local problem we had hitherto been describing. We split the Causes into local and systematic branches, and added this to the systematic branch. However, noting that there is currently only one other type of systematic error Causes, this may not be the best way to organize these concepts.

The Causes branch of our taxonomy is likely underdeveloped. Many more possible causes of errors are likely to exist than what we have included. We are interested to see how the description of additional causes might create or reform the structure of the Cause branch, and how they might lead to new ideas for detecting and preventing problems.

Fundamental errors in the modelling schemas.

Rector et al. use this to describe the well-known problems with inheritance and inference that result from misuse of the legacy SEP triple schema in Anatomy. This category poses an interesting question for our taxonomy. If an ontology is designed with certain semantic functions, but users repeatedly fail to model terms correctly within those semantics, where do the errors originate? We assert that the errors originate with the user, and thus have categorized this under *User problems: misuse of semantics* as *Incorrect use of SEP triples*. However, the best way to avoid this type of user mistake may very well be to change the semantics of the ontology to be easier to use and understand. The Anatomy branch is currently being revised to eliminate the use of SEP triples.

Some thought was given to problems that appear to be more complex or distant than the simple and immediate error set proposed in this taxonomy. Rector et al. note that "A set of definitions and axioms, however individually plausible, contains an error if it leads to an erroneous inference, as judged by domain experts." [2] This author believes that such errors can still be described by this taxonomy.

For example, Rector et al. found *Hypertension* is inferred to be under both *Finding* and *Disorder of Soft Tissue*, several levels away in the hierarchy. *Hypertension* should not be under anything related to *Soft tissue*, so they investigated the source of the problem, which was not immediately apparent. Although the problem was *detected* by a disconnection between distant concepts, the *error* was actually immediate.

> "Tracing the cause of both inferences to their root, it was found that they both followed from the axiom that the *site* of *Hypertensive disorder* was *some Artery* and that *Arteries* were classed as *Soft tissues*. Together, these axioms led to the unwanted inferences…. Therefore, the axiom was changed so that *Hypertension* was *sited* simply in the *Cardiovascular system."*

In other words, the problem was ultimately reduced to an immediate problem: *Hypertension* had an "attribute too specific" error, which was corrected by replacing the incorrect attribute. Rector notes that it looks initially reasonable to have *Artery* as the *finding_site* of *Hypertension*, even though *Cardiovascular system* is found to be more accurate. An inferential error may be difficult to detect if the error is not apparent in the immediate relationship. We do not contest this assertion. We merely suggest that such inferred errors can ultimately be described by the proposed taxonomy, even if the taxonomy does not outline the means to detect them.

However, for the purposes of the submission request system, it may be prudent to develop a means for submitting errors that are not dissected down to their ultimate origin. That is, if a submitter notices that *Hypertension* should not be found under *Soft Tissue* disorders, but has not deduced the originating error, it would still certainly be worthwhile for them to submit the inferential error for examination.

Results and Discussion: NJIT Error Analysis

All numbers from the analysis of these error reports are available in Table 2. The original error report data is available in a related Access file. Within the 108 reports examined, 135 individual errors were reported. Of these 119 were actual errors, while 16 were invalidated as actually correct. An additional 26 errors were found in the submitted suggestions to fix errors (in other words, the suggestions were wrong). These numbers are not necessarily representative of the actual distribution of errors in SNOMED CT, nor even of the entire set of error submission. However, they are informative nonetheless, and possible trends merit further investigation.

Since this sample is drawn from the specimen tree, many of these errors relate to the anatomy tree. Problems with the anatomy hierarchy are widespread and well-known, and an extensive revision of this tree is already planned. The point of analyzing this data is not to uncover further error or to draw attention to known error, but to show how errors may be categorized, described, and detected.

Table 2: NJIT Errors and Causes

					Problem Concept		Incorrect Concept		Related Word		Incom-pletion		None		Other		
Problems	**Submitted Instances**	**Actual Instances**	**In Suggestions**		**Concepts**	**Instances**	**Concepts**	**Instances**	**Concepts**	**Instances**	**Concepts**	**Instances**	**Concepts**	**Instances**	**Concepts**	**Instances**	**Ancestral Origin of Error**
Parent too general	21	21	1		5	5	0	0	0	0	0	0	16	16	0	0	2
Parent too specific	12	12	1		5	5	4	6	0	0	0	0	1	1	1	1	9
Parent unrelated	15	12	10		7	7	5	5	3	3	0	0	0	0	0	0	1
Parent missing	22	11	0		2	2	1	1	0	0	0	0	19	19	0	0	0
Attribute missing	58	58	0		0	0	0	0	0	0	58	58	0	0	0	0	0
Attribute too general	0	0	12		0	0	0	0	0	0	0	0	12	12	0	0	0
Attribute too specific	0	0	1		0	0	0	0	0	0	0	0	1	1	0	0	0
Attribute unrelated	3	1	0		1	1	0	0	1	1	0	0	0	0	1	1	0
Missing description	1	1	0		0	0	0	0	0	0	0	0	1	1	0	0	0
Redundancy	2	2	0		2	2	0	0	0	0	0	0	2	2	0	0	0
Disjunctive Aggregation	1	1	0		0	0	0	0	0	0	0	0	1	1	0	0	0
Term is out of scope	0	0	1		0	0	0	0	0	0	0	0	1	1	0	0	0
Totals*	135	119	26		19	22	9	12	3	4	58	58	49	54	2	2	0

Key to Table 2	
Submitted Instances	These are the errors as submitted by users
Actual Instances	These are the errors as determined by reviewers
In suggestions	These are errors that would be introduced by submissions intended to correct errors
Concepts	This is the number of concepts that have at least one of this type of error associated with the request
Instances	This is the number of actual instances of each type of error. There may be multiple instances of a single type of error associated with a concept, thus the difference between concepts and instances.
Ancestral Origin of Error	This is the number of errors of this type that appear to have their ultimate cause in some type of problem elsewhere in the SNOMED hierarchy.

Highlights from Errors and Causes

Missing Attributes and Incompletion

By far, the most common error reported in the sample was a missing attribute, as shown in Figure 1. Every instance of it was caused by what we called Incompletion, as shown in Figure 2. Incompletion means that a relationship needed to fully specify a concept is absent. In this sample, which is drawn entirely from the specimen tree, these missing attributes are generally a source topography for the specimen, a substance of the specimen, or a laterality for a specimen that specifies a side in the name.

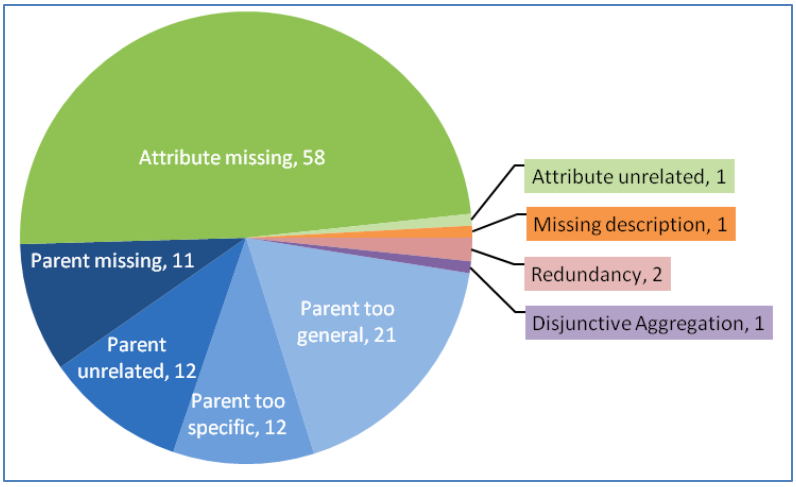


Figure 1: Errors in NJIT Sample

Source topography and substance are defining attributes for all specimen concepts. If this sample is at all indicative, these are likely to be missing from a great many more. However, it should be fairly easy to detect their absence computationally. Unfortunately, errors detected would then require manual review. Similarly, laterality can be used with any term where a side is indicated, though it is not a required as a defining attribute for specimen. In this case, the problem should be both simple to detect, and easy to correct computationally. Although these particular relationships may be specific to specimen, there other concept trees have similar requirements for full specification. Therefore, computational detection of this simple type of error should be possible elsewhere, as well.

Problematic Conceptualizations

Incorrect conceptualizations and problematic conceptualization, where the concept is misunderstood by either the ontology creator or the user, give rise to a variety of errors. Together, *Problem Concept* and *Incorrect Concept* represent the cause of about a quarter of the errors in the sample. A significant portion of erroneous suggestions result from these kinds of misunderstandings, as well.

Some of this confusion may be unavoidable. For example, many of the errors in this sample are related to the term “soft tissue.” Soft tissue has varying definitions within medical communities. In some, soft tissue includes all non-bony tissues. However, other medical professionals use the more limited definition SNOMED CT has adopted, which does not encompass glands and organs. Other groups may exclude additional tissues, like nerves, that SNOMED CT includes.

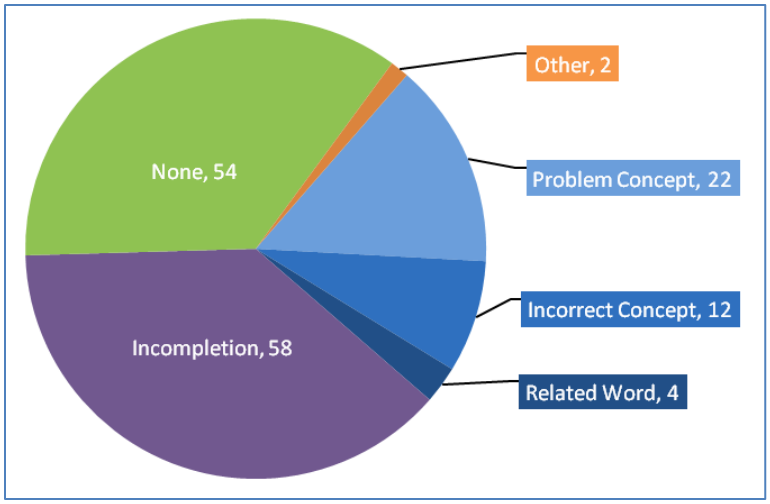


Figure 2: Causes of Errors

No matter what definition is used in SNOMED CT, the result would be counter to the expectations of some users.

In other cases, the confusion may be resolvable. For example, the submitter and reviewer disagree on the definition of synovial fluid specimen, as to whether it includes pathological states of joint fluid or only the healthy state. Although this author is not qualified to make a decision on this topic, additional sources do seem to confirm the more inclusive definition. Therefore, some of the errors marked as *Problem Concept* may actually be considered *Incorrect Concept* when reviewed by additional experts. This suggests it may be useful to distinguish between concepts with resolvable consensus definitions, and concepts with necessarily non-consensus definitions. In either case, providing textual definitions in cases where a concept is potentially confusing would help eliminate confusion over whether the logical definition is correct.

Table 3: Causes Key

Problem Concept	Problem with conceptualization of whole concept, component of concept, or related concept in question. Concept may have been incorrectly conceptualized, or it may be difficult to find a consensus definition for it.
Incorrect Concept	Concept has definitely been incorrectly conceptualized
Related word	Problem with the logical definition originates from a word within the concept related to the error
Incompletion	Failure to specify some aspect of a concept required by the editorial guidelines to define the class. For samples, this is often topography, substance, or laterality.
Ancestor	Problem originates higher in the hierarchy elsewhere than in the immediate relationships
None	No apparent cause for error. These are likely to be "simple mistakes"--selection of a concept that is too general, selection of a concept before a better option was added to the vocabulary, etc.

Rector and Iannone similarly observed a suite of error resulting from ambiguity in the use of the modifiers “acute” and “chronic.” [3] They conclude that the ambiguity reflects the different uses of these words in patient care versus pathology. In making a decision as to which definitions should be preferred in SNOMED CT, they suggest patient care definitions are more likely to have direct consequences to health.

No Apparent Cause

Over a third of the errors in the dataset have no apparent cause, and are marked *None* in Figure 2. These are likely what Rector et al. calls “Simple Errors,” caused by simple carelessness, lack of

thoroughness, or lack of familiarity of with the ontology [2]. However, it possible that analysis of a larger dataset might reveal additional patterns of error belied by this category.

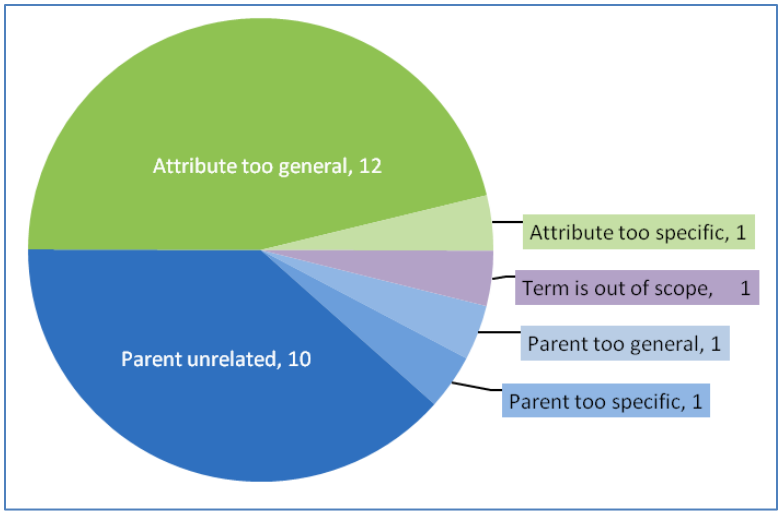


Figure 3: Errors in Suggestions

Suggestion Too General

Of the errors in suggestions, half are suggestions that are too general—i.e., a direct descendant of the suggested term would be a better, more specific choice. Most of these are attributes, with only one being a parent term. For example, *body tissue material (substance)* [413675001] is suggested as a parent to *specimen from sympathetic nerve ganglion (specimen)*. However, *Nervous system material (substance)* [277297006], which is a child concept of *body tissue material*, would be a more appropriate parent. While these submissions are not wrong in the strictest sense, it is definitely preferable for the more specific terms to complete the hierarchy. This kind of error could probably be prevented in individual submissions by prompting submitters to check the children of the term they are suggesting. Immediate feedback that asked something such as, "Are any of these children terms better choices?" could make it easy to select the correct term, and could be repeated iteratively, if necessary to come to the most specific choice. However, preventing this type of error in batch submissions of the type NJIT submitted would be impossible.

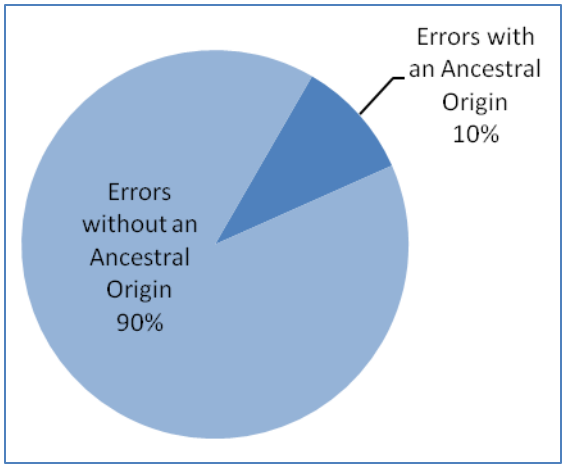


Figure 4: Ancestral Errors

Ancestral Errors

Of the 120 errors in the dataset examined, 13 errors were rooted in a deeper problem with the ancestral hierarchy. Ancestral problems are of special interest, because of their potential reach in creating inferential errors in many additional areas of the hierarchy. Therefore, we have given them close analysis in this report. All ancestral errors in our dataset gave rise to errors in parental relationships. Two concepts from this set also had attribute errors, but since these did not arise from the ancestral problems, those attribute errors are not included in this analysis.

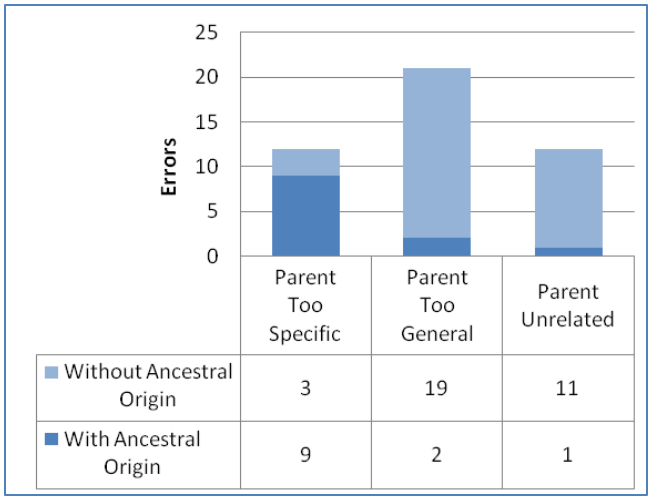


Figure 5: Ancestral Errors by Type

Parent Too Specific:

Six of these nine errors pertain to the incorrect conceptualization of a single part of the anatomy, the ampulla of Vater. Another error results from a potentially incorrect conceptualization of another part of the anatomy, the placenta. Two of the other errors arise from confusion over the conceptual differences between excision, incision, and resection. These could be incorrect conceptualizations or simply problematic concepts where a consensus conceptualization does not exist between medical specialties.

In the case of the errors related to anatomy, the specimen tree is not the correct place to ultimately alter the modeling. In the case of terms dealing with surgical techniques, alterations at this level may or may not be correct, depending on the ultimate definitions of excision, incision, and resection.

It is interesting to note that in the remaining five instances of *Parent too specific*, two were also caused by conceptualizations that are potentially incorrect, or at minimum problematic. Since these types of errors appear to often cause problems elsewhere in the ontology, our tentative conclusion is that *Parent too specific* is an error that strongly signals existing or potential problems with the definitions of concepts that are likely to affect other areas of the hierarchy. Identifying what part of the concept caused the error could allow any related terms elsewhere in the ontology be investigated for additional errors. Rector et al. have successfully used a similar approach. Starting from the recognition of an individual problem with *hypertensive retinopathy*, they identified and corrected problems in four additional classes of hypertensive disorders. [3] "Parent too specific" errors could be used as a launching point for employing methodology similar to the procedure followed by Rector et al.

Parent Too General:

The first reported error suggests removing parent *specimen from bone* (430268003) from *iliac crest bone marrow sample* [271515009]. The submitter points out that this is a bone marrow specimen, not a bone specimen. The key problem in this case appears to be the conceptualization of bone marrow, which is represented in SNOMED CT primarily as part of the hemopoetic system and not as part of bone. While this conceptualization is arguably valid, it is also not surprising that it may engender confusion and error. (Thus, the suggestion, which is classed here as a *Parent too general* error, is actually a *Parent unrelated* error.) In fact, tracing through the anatomy hierarchy, we find that *iliac crest marrow structure* [313227001] is represented as part

of bone via the following pathway: parent of *Iliac crest structure* [29850006], parent of *ilium part* [119539000], parent of *bone part* [119186007], *Structure of bone* [421663001]. This is not the case for several other types of marrow examined, but a comprehensive check was not performed.

The second suggestion removes the parent *specimen from soft tissue obtained by fine needle aspiration biopsy* [441810001] from *specimen from sublingual gland obtained by fine needle aspiration biopsy* [441876003]. The submitter notes, "The removed parent should be attached much higher in the hierarchy." The key problem here appears to again be the problematic understanding of "soft tissue" in SNOMED CT. Since glands are not included in SNOMED CT soft tissues, the parent is indeed incorrect. However, it should not appear higher in the ancestry for this term. The original error seems likely to have resulted from confusion over what soft tissue encompassed. The recent recommendation to remove it actually reflects the same confusion, implying that higher in the hierarchy, soft tissue specimen should be a parent to gland specimen.

In both cases, the ancestral problem is related to a concept that is poorly conceptualized or, at best, ambiguous to users. This appears to be the origin of three additional *Parent too general* errors without ancestral origin, so this state of conceptual ambiguity does not alone indicate a deeper ancestral problem. However, these five still represent a minority of the 20 errors in this category overall, so incorrect or questionable concepts could be used as a signal to further investigate this type of error.

Parent Unrelated:

Here, *specimen from blood product* [119300005] has the apparently unrelated parent *device specimen*[127454002] . A replacement parent of *drug specimen* [119319000] is suggested. The reviewer notes in response, "Currently, blood products are listed as devices; this request would require a remodeling of the device specimen hierarchy.… I also do not agree with the proposed parent of drug specimen, especially when the blood product is homologous." One key problem here is a conceptualization of blood products as devices that is questionable at best, possibly incorrect, and certainly puzzling to users.

The *device specimen* concept appears to encompass two conceptually different entities: medical objects that imply collection of a biological specimen (*urine collection bag submitted as specimen* [439628004]) versus biological specimens obtained from such medical objects either implicitly or explicitly (*urine specimen obtained from urine collection bag* [446306009]). Device specimen might be an ambiguous concept that should be separated into distinct ideas. Additionally, all body fluids must presumably be collected via some medical device, but not all are subject to modeling based on this fact. Cerebrospinal fluid specimen, for example, is not modeled as dependent on any devices. The ontology appears to lack a consistent treatment of fluids and the collection devices for them.

It is also certainly incorrect to suggest that *drug specimen* would be the parent *of specimen from a blood product*, when *drug specimen* is only one type of specimen that could be derived from blood. However, not all drug specimens need be derived from blood, so *drug specimen* cannot be a child concept of *specimen from blood product*, either. Instead, both should be parents to a more specific concept like *drug specimen derived from blood product*. This suggestion perhaps represents a mistake from attempting to "force" the existing concepts to do the work of the missing concept. Although this is the only such example in the sample, it is possible that concepts that appear to have unrelated parents or parents that are too specific may predict missing concepts of this type.

Since the underlying problem in all cases of ancestral errors in this sample appeared to be a concept that was incorrectly or confusingly conceptualized, the character of the ancestral problems that can be identified from this type of analysis may be limited.

Recommendations

Confirm findings with additional data

The exploratory samples analyzed for this report were small and non-random. Additionally, the same sample coded with the taxonomy was used to develop the taxonomy. A larger, more representative sample must be analyzed to confirm or refute the findings of this report and usefulness and comprehensiveness of the error taxonomy in describing errors. Alternatively, a sample targeted towards just one type of error and its implications could be performed. For example, requests to remove incorrect parents could be analyzed to further explore the possibility of using them to identify ancestral origins of these errors.

We have also suggested that a number of errors that may indicate other similar. Perhaps most importantly, *Parent too specific* errors may indicate underlying ancestral errors that have propagated elsewhere in the hierarchy as well, potentially engendering bad fixes at the level of individual concepts. Additional possible patterns were observed in this dataset. *Parent too general* , *Parent missing,* and *Attribute missing* errors may predict similar errors in sibling concepts. Also, where the cause of error was *Related word,* a related word in the concept leading to an incorrect association, the related word may have led to a similar problem elsewhere. The validity of these possible patterns should be further investigated.

Reduce Conceptual Ambiguity

Track concepts that engender confusion

As mentioned above, soft tissue is a term that does not appear to have a consensus definition among medical professionals, and therefore it is likely that the SNOMED CT definition will remain a source of error reports no matter what definition is used. Resection may prove to be another such concept. Adopting some way to track concepts like these as necessarily problematic

may help make it easier to respond appropriately to these requests, or even prevent their submission.

Provide textual definitions for concepts

Logical definitions are not easy for all users to understand. Providing textual descriptions for SNOMED CT concepts could help clarify the intent and scope of confusing terms. We recommend prioritizing concepts that are noted as confusing or have a necessarily non-consensus definition, like those noted above. For example, concepts marked as problematic could automatically display their textual definitions if a user attempted to submitted a related request.

Revise the SNOMED CT Request Submission System

Improve Usability

The current request system lacks clarity and consistency in its language, and has insufficient documentation for naïve users. We have recommended a fairly complete overhaul of the menu options for the submission system below, so we will not detail shortcomings of the current language and navigation options here. Testing of the current system identified many additional improvements to usability that are not fully relevant to this report, not being concerned particularly with the reporting of errors. Detailed feedback has been submitted to the developers of the request system that could improve the system in its current form. This feedback is included in Appendix B. However, whatever system is ultimately adopted, care should be given to using consistent, unambiguous language and providing adequate documentation.

Include and Organize Essential System Tasks

Because the complete list of Actions includes sweeping changes normal users would not need to effect, such as retiring entire relationship terms, the request submission system only needs to handle a subset of the Action branch of the taxonomy. These are represented below.

Actions (Submission system subset)

- Create term
    - Create concept
    - Create description (including synonyms)
- Retire Term
    - Retire concept
    - Retire description
- Modify Term
    - Edit dependent features
        - Edit lexically
        - Edit concept features
            - Edit FullySpecifiedName
        - Edit description features

            - Edit DescriptionType
    - Add relationship to concepts
        - Add parent
        - Add attribute
    - Remove relationship from concepts
        - Remove parent
        - Remove attribute
    - Add descriptions to concepts
    - Remove description from concepts

It may be worthwhile to let users make additional requests in free text, as well, to allow them to report systematic causes of error like lack of normalization, or to report an error that they are unsure how to correct. Although it is important to have each of these actions available in the system, obviously, each does not require its own form. Actions and forms could be organized around one of two principles, either by task or by concept.

To organize by task, the menu would reflect a similar, but simplified organization to the current implementation. Several related options could be implemented on a single form. For example, it's unlikely that a user would want to create a new synonym without adding it to some existing concept, so these could be two available options on a single form. Major menu options could include:

- Create new concept
- Retire existing concept
- Edit relationships
- Edit synonyms and other descriptions

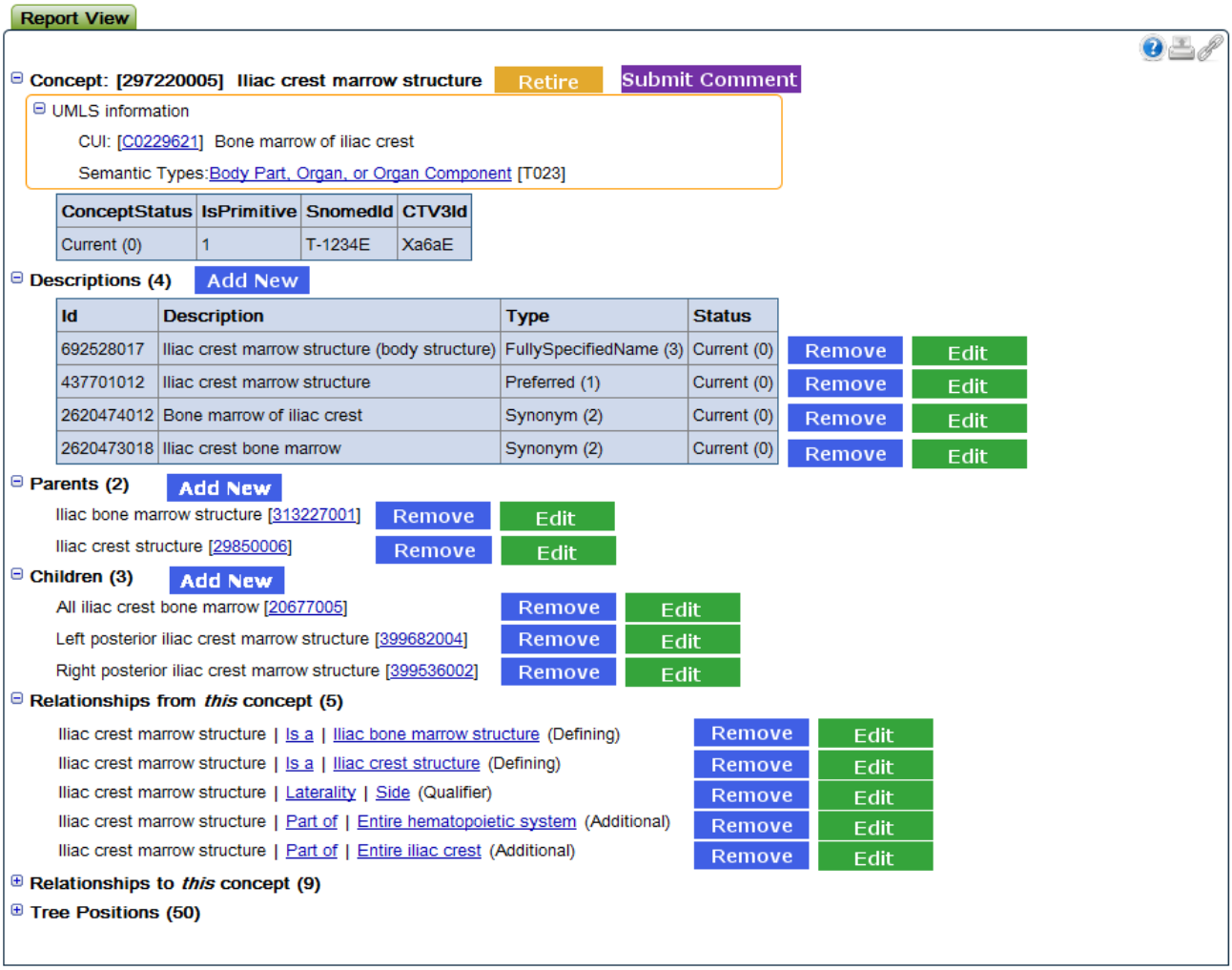
The full suite of available actions would then be on four main forms:

- Create new concept
- Retire existing concept
- Edit relationships
    - Edit existing relationship
    - Remove existing relationship
        - Retire existing concept
    - Add new relationship (add parent, attribute, etc.)
        - Create new concept
- Edit synonyms and other descriptions
    - Edit existing synonym or description
    - Remove existing synonym or description
        - Retire synonym or description
    - Add new synonym or description to a concept

- Create new synonym or description

However, as demonstrated by the NJIT dataset, a significant proportion of concepts requiring revision will have multiple error types and will thus require multiple corrective actions. It may be more efficient for users, therefore, if this system is implemented around concepts and not around tasks. In this case, the main interface might be built around a familiar browser like the UMLS SNOMED CT browser, with options given to alter the erroneous information or add new. Subforms would be required to open dynamically for complex editing and to prompt users to comment on the reasons for their requests. Figure 6 contains a simple mockup of this basic idea for reference.

Figure 6: Browser-Based Request System Mockup



Prompt the Submitter

Submitters could be guided and prompted to provide higher quality suggestions, additional information, and additional review during the submission process.

Ask for the Error and Cause

We had hoped to be able to derive the type of Error from the Actions the user wishes to take, but analysis shows this to be too problematic. First, many concepts have multiple unrelated or related Errors, and trying to tease apart the Errors based only on what Actions were taken would be impossible in these cases. Additionally, even if there is only a single Error, the submitter may not know exactly how to correct it. For example, they might remove a parent that is too general, without offering a replacement parent as expected. In this case, the requested Action would suggest the Error is *Parent Unrelated*, not *Parent Too General.* Thus, it ultimately seems impractical to attempt to derive the Error type from the Actions.

Therefore, it would be useful if submitters would specify and Error type and a Cause for their request. Knowing this information would be helpful for a variety of reasons. Obviously, it would helpful to the reviewer's analysis and decision-making on the request. For example, since it appears that some problems like *Parent too specific* seem to be more likely to originate from an ancestral problem, this would flag those concepts for especially thorough investigation. Additionally, it would allow us to prompt the user to make more complete suggestions. For example, in the case of a *Parent too general* error, we could prompt the user to not only remove the incorrect term, but select a replacement.

Give Instant Feedback

Instant feedback could be used to improve the quality of individual submissions. For example, the most common problem found in the NJIT suggestions was a parent or attribute that is too general. If the child concepts of the original suggestion could be instantly displayed with a prompt like, "Would any of these child concepts be more suitable?" these errors could be dramatically reduced. Instant feedback could also be employed to allow the user to see if they have made a mistake. This seems especially likely when creating new relationships. It is easy to accidentally reverse the directionality of a relationship, so an immediate rephrasing of the request as a triple or a visual representation of the new relationship would likely prevent errors. Obviously, this type of feedback would not be practical for submissions that are uploaded in large batches, where the user does not interact with the system for each error reported. However, it might be worthwhile for individual submissions.

Suggest Additional Review

While the user is already immersed in the system and looking at a specific area, we might be able to encourage them to review additional related concepts. For example, we could encourage them to examine sibling concepts for related error with prompts like, "Do any of these sibling concepts also need this new parent?" If specific predictive patterns can be established for errors that are generally propagated across a set of terms, more targeted feedback systems could be incorporated to make the most efficient use of the human brainpower available for reviewing SNOMED CT. Since human review is likely to be the most effective, important source of evaluation for the foreseeable future, getting the most out of human reviewers should be a priority.

References

[1] David Robinson, "SNOMED CT Quality Review Phase 2 Draft Methodology Criteria, Tooling and Usage," NHS-PROG-RPT-99999-020-SNOMED CT , Version 0.3, Jan. 2011.

[2] A. Rector *et al*., "Quality Assurance of the Content of a Large DL-based Terminology using Mixed Lexical and Semantic Criteria: Experience with SNOMED CT," *Proceedings of the sixth international conference on Knowledge capture (KCAP 2011)*, pp. 57-64.

[3] A. Rector, S. Brandt, and T. Schneider, "Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications," *JAMIA*, vol. 18, no. 4, pp. 432-440, Apr. 2011.

[4] A. Rector and L. Iannone, "Lexically Suggest, Logically Define: Quality Assurance of the Use of Qualifiers and Expected Results of Post-Coordination in SNOMED CT," *JBI*, submitted for publication.

Appendices

Appendix A

SNOMED CT Error Taxonomy

Errors

Semantic errors

- Incomplete logical definition
  - Missing parent
    - *Ex: "cerebrospinal fluid sample (specimen)" [258450006] is missing parent "specimen from central nervous system (specimen)" [399436000]*
  - Missing attribute/value (laterality, etc.)
    - *Ex: "joint fluid specimen (specimen)" 431361003 missing specimen source topography "articular space (body structure)"[36668004]*
  - Missing full definition despite the presence of necessary and sufficient conditions.
    - *Heart disease should be fully defined with "Any disease with site heart." However, it is instead primitive with the provision that it includes "Some diseases with site of heart."*
- Incorrect logical definition or conceptualization
  - Incorrect parent
    - Parent too general

        - *Ex: “meconium specimen (specimen)” [119340004] has parent too general: ”body fluid sample (specimen)” [309051001]. Could be replaced by “fecal fluid sample (specimen)*” [*258457009]*
    - Parent too specific
        - *Ex: “frenulum of penis swab (specimen)” [258516004] has parent too specific "glans penis swab (specimen)" [258512002]. Could be replaced by "penis swab (specimen)" [258510005]*
    - Parent unrelated
        - *Ex: “specimen from ampulla of Vater (specimen)” [127467004] has unrelated parent “specimen from pancreas (specimen)”[127469001]. (“The ampulla of Vater is definitely not part of the Pancreas.” –NJIT submitter)*
    - Parent redundant
        - *This occurs when a parent that is given explicitly can also be inferred.*
  - Incorrect attribute
    - Attribute too general
        - *Ex: “specimen from axillary lymph node obtained by fine needle aspiration biopsy (specimen)” [441694006]. NJIT has recommended a too general specimen substance "body tissue material (substance)" [413675001]. Alternate recommendation for substance is “Lymphatic material (substance)” [289967009]*
    - Attribute too specific
        - *Ex: “oral secretion sample (specimen)” [258558001]. NJIT has recommended a too specific specimen source topography "Entire oral cavity (body structure)" [181220002]. Alternate recommendation for source topology is “oral cavity structure (body structure)” [74262004].*
    - Attribute unrelated
        - *Ex: “tissue specimen obtained from gallbladder by cholecystectomy with partial hepatectomy (specimen)” [369617007] has unrelated source topography “Liver structure (body structure)” [10200004]*
    - Attribute redundant
        - *This occurs when an attribute that is given explicitly can also be inferred.*
- Disjunctive Aggregation
  - *Ex: “lymph node from sentinel lymph node dissection and axillary dissection (specimen)” [384744003]. It is unclear from which part of the surgery the lymph node specimen originates.*

- Redundant concept
    - *Ex: "specimen from appendix and right colon obtained by appendectomy and right hemicolectomy (specimen)"[422991009] is a duplicate of its parent "Specimen from appendix and colon obtained by appendectomy and right colectomy (specimen)" [423696009].*
- Meaninglessness outside of classification
    - *Concepts that contain "not otherwise specified", "other", or any other categorization described only by exclusion from related categories.*
- Concept missing from vocabulary
    - *Ex: The" apophyseal ring" or "ring apophysis" appears to be a missing concept in SNOMED CT, originating in RADLEX. It would presumably be a child of "apophysis (body structure)" [65000005]*
- Concept out of scope
    - The concept is not within the scope of SNOMED CT. Examples might include non-clinical scientific terms or slang.

Lexical errors

- Missing synonym
    - *Ex: "transfusional siderosis" (from Radlex) is a synonym for "transfusion hemosiderosis (disorder)" [69281008] but is not currently given as a synonym.*
- Incorrect description
    - Incorrect synonym
        - Non-equivalent synonym
            - *A given synonym is not actually equivalent to the related concept*
        - Non-clinical synonym
            - *The synonym is not suitable for clinical use*
    - Incorrect preferred term
        - *A better preferred term for the concept exists than the preferred term currently given*
    - Incorrect fully specified name
        - *The fully specified name does not represent the concept*
    - (Incorrect supplemental text description)
        - *The supplemental text description is an inaccurate definition of the concept*
- Incorrect Spelling
- Incorrect Punctuation
- Incorrect Pluralization
- Incorrect Capitalization
- Incorrect Abbreviation
- Incorrect Spacing

Actions

- Create term
    - Create concept
    - Create description (including synonyms)
    - Create relationship
- Retire Term
    - Retire concept
    - Retire description
    - Retire relationship
- Modify Term
    - Edit dependent features
        - Edit lexically
            - Edit concept features
                - Edit CTV3ID
                - Edit IsPrimitive
            - Edit relationship features
                - Edit characteristic type
                - Edit refinability
                - Edit Relationship Group
            - Edit description features
                - Edit Capitalization status
                - Edit LanguageCode
    - Add relationship to concepts
        - Add parent
        - Add attribute
    - Remove relationship from concepts
        - Remove parent
        - Remove attribute
    - Add descriptions to concepts
        - Add synonym
        - Add fully specified name
        - Add preferred term
        - (Add supplemental text description)
    - Remove description from concepts
        - Remove synonym
        - Remove fully specified name
        - Remove preferred name
        - (Remove supplemental text description)

Causes

- None apparent
- Local Causes
    - Problematic Conceptualization
        - Incorrect Conceptualization
            - Over-Literal Definition
        - Confusing Conceptualization
            - Non-consensus conceptualization
            - Ambiguous conceptualization
            - Non-intuitive naming convention
    - Inadequate concept model
    - User Problems
        - Misuse of semantics
            - Incorrect use of assertion
            - Incorrect use of SEP triples
        - Misuse of related word in concept
        - Local over-assertion
- Systematic Causes
    - Lack of normalization
    - Inadequate

Appendix B

Usability Feedback on the SNOMED CT Beta Request System

Language

The current menu heading "change parent" suggests that you will be able to not only add a parent, but remove one as well, which is not the case on the form. However, the option "Change relationship" lets you *modify* an existing relationship, but not add one or remove one. This is quite a different meaning of "change." A menu option word like "change" should always indicate the same actions to a user.

Additionally, the menu options "change" and "retire" are both verbs. A third menu option, "new" (like "new concept") is not a verb, and should perhaps be replaced with something like "create." Use language consistently.

The granularity of the options available from the main menu is not consistent. For example, "Change parent," which allows you to add a parent, is technically a subset of the "new relationship" category of actions. Both are in the menu. An argument can be made for breaking the granular architecture in order to make a common task easier for a user to get to. However, adding an attribute might be a more common task than adding a parent.—I'm not sure adding a parent needs its own category. However, if it does, perhaps "new relationship" should be "other new relationship," or "add other relationship" to be clearer?

Also on the new relationship form, I assume that the “relationship type” is meant to be the new, suggested relationship type and not the existing one. However, the text of the roll-over box seems to indicate that it should be the existing relationship. This could be clarified.

Additionally, “synonym” and “description” seem to be used interchangeably in the current system. Synonyms represent one subtype of description, and the correct words should be chosen where appropriate.

Is “retire concept” for removing a concept entirely from active use, but “retire relationship” is just to remove a specific instance of a relationship between two concepts, not to retire an entire relationship term from the vocabulary? Is there an option for this? Again, the language here is a bit confusing.

New request forms

As someone not really well-acquainted with the structure of Snomed, I found some of the boxes here fairly confusing. The roll-over information was not always helpful. I know a naïve user like me is not your main audience, but I find it possible that other people without an expert level of familiarity with the structure of the vocab might make requests. There may be researchers using Snomed to tag data, for example, but not acquainted with all the ins and outs of what characteristic types are and so on. At the level of understanding I have, I would like more documentation on what should be added to each box.

For “Topic,” for example, I think, but am not sure, that these are from a controlled vocabulary. In that case, it would be really nice if additional boxes could help with filling in the vocabulary. Topic, for instance, could have a drop list. If that list is very long (I’m not sure how many topics there are in Snomed), feature like autoskip or autocomplete could help find the correct term in real time. Similarly, a browse box that let you search for the concept ID or term you are looking for, and could autofill the appropriate related boxes (ID should also autofill the term, for example) would be really helpful and probably cut down on errors in submission. I know that you can go to an external browser for this, but it would be much more convenient and less error-prone to have it internally.

Under “Change Relationship”, the roll-over instructions for Characteristic Type and Refinability do not fit on my monitor, which is 20in. These need to wrap the text on two lines, or some other system for conveying this type of information needs to be applied (the little question mark button with a pop-up box is very common, for example).

IDs should probably employ an input mask to force correct ID format. I put in all kinds of values, and it accepted everything, including 0 and alphabetic letters.

I have a suspicion that the forms that deal with relationships will be most likely to generate mistakes, because of the directionality of the relationship. You could add in feature that would

write out the triple after you have entered it (butterfly band-aid is_a band-aid) so that people get instant feedback on what they have entered and can see right away if they did it wrong.

My Requests Screen

Upon submission of a draft batch, feedback that your batch has been successfully submitted would be nice.

What is the difference between accepted and approved requests?