SNOMED CT Revisions and Coded Data Repositories: When to Upgrade?

Journal:	*AMIA 2011 Annual Symposium*
Manuscript ID:	AMIA-0745-A2011.R1
Manuscript Type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Ceusters, Werner; NYS Center of Excellence in Bioinformatics & Life Sciences, Ontology Research Group

SNOMED CT Revisions and Coded Data Repositories: When to Upgrade?

Werner Ceusters, MD[1]
[1]New York State Center of Excellence in Bioinformatics, University at Buffalo, Buffalo, NY

Abstract

SNOMED CT is gaining momentum in its acceptance and operational application as a reference terminology in electronic health systems. Because it is revised every six months, organizations using SNOMED CT might feel a need to ensure that their systems are synchronized with these revisions. It has been shown that for certain sorts of applications migration to a new version is a labor-intensive process. Here two indicators – the evolution of the global information content of an ontology over consecutive versions, and the perseverance of suspicious events – are proposed to assess whether it is worthwhile upgrading when a new version is released. The indicators can be computed automatically when a new version is released and are statistically unrelated. Trend breaks in their evolution are suggestive for the possible benefit of an upgrade and their predictive power correlates well with the retrospective realism-based quality metric which forms the basis of Evolutionary Terminology Auditing.

Introduction

SNOMED CT is a clinical reference terminology for annotating patient data designed to enable electronic clinical decision support, disease screening and enhanced patient safety [1]. It is structured around a taxonomy of what are called *concepts* which are associated with a variable number of assertions concerning the *relationships* of such concepts either to other concepts or to *terms* linked to the concepts by means of *descriptions*. Whereas the descriptions provide a vocabulary for talking about the concepts (or what might be instances thereof when the vocabulary is used to annotate patient data), the concepts and relationships themselves are supposed to be a representation of what exists in reality and is of relevance for certain purposes in biomedicine.

The content of SNOMED CT evolves with each release. Types of changes involving the core components include the addition or deletion – including replacements – of concepts, descriptions, and relationships. A history mechanism keeps track of these changes over time, thereby adhering to one of the well-known requirements for terminology management proposed by Cimino [2]. Figure 1 shows, as an example, the impact of taxonomy changes on the concept '*Cell phenotyping performed (situation)*' for all versions from January 2002 to July 2010. The arrows represent the relationships as they are found in the relationships table and are labeled – to the right of or above the arrow – with the preferred name of the relationship and the period(s) during which it was stated to hold; a label such as '0501-1007' thus signifies that a relationship was introduced in the January 2005 version and has been present to July 2010. The arrows are further color coded for quick visualization of their history: red, green and blue mean, respectively, that a relationship is found *prior to* the previous version, *in* the previous version, and in the *latest* version.

As can be seen in Figure 1, changes have been quite dramatic over time. This raises several questions concerning the impact these changes have on data collections which are coded in terms of – usually a small subset of – SNOMED CT concepts. Sensible questions are, for instance, (1) when it is worthwhile to use a new version, since revisions made may be wholly outside the scope of the data collected, (2) whether analyses performed using an earlier version are rendered meaningless because of the inactivation of concepts in later versions, and (3) whether a new version contains more or less knowledge than its predecessor or is a mere reformulation of the same amount of knowledge. The purpose of the work described in this communication was to establish whether answers to such questions can be found, and what would be possible strategies to find reliable answers in operational environments in which research on such issues is not part of the core activities.

Background

SNOMED CT has primarily been researched in terms of (1) the coverage that it provides to support coding in specific domains [3-4], (2) the reliability and validity of such coding efforts [5-6], and (3) its ontological coherence and consistency [7-9]. Less thoroughly studied is how SNOMED CT evolves over time: while some focus their efforts on descriptive statistics involving the mere appearance and disappearance of concepts, terms and relationships [10], the *Evolutionary Terminology Auditing* method attempts to translate such changes into quality measures indicating (1) how much a new version of a terminology is better than any previous version and (2) to

what degree terminology changes reflect changes in the underlying domain (for example appearance of new diseases) or in the terminology authors' understanding thereof [11].

To the best of our knowledge, only Wade and Rosenbloom have thus far addressed the impact of SNOMED CT's evolution on operational applications, with the conclusion that '*While the efforts of each subsequent SNOMED CT version aim for continual improvement, changes made to its core structure and post-coordination guidelines make it more difficult to migrate proprietary data to this reference standard*' [12]. That this issue thus far has not received the attention that it deserves can be explained by the rather limited number of actual implementations, a situation that probably will change dramatically in the near future [13].

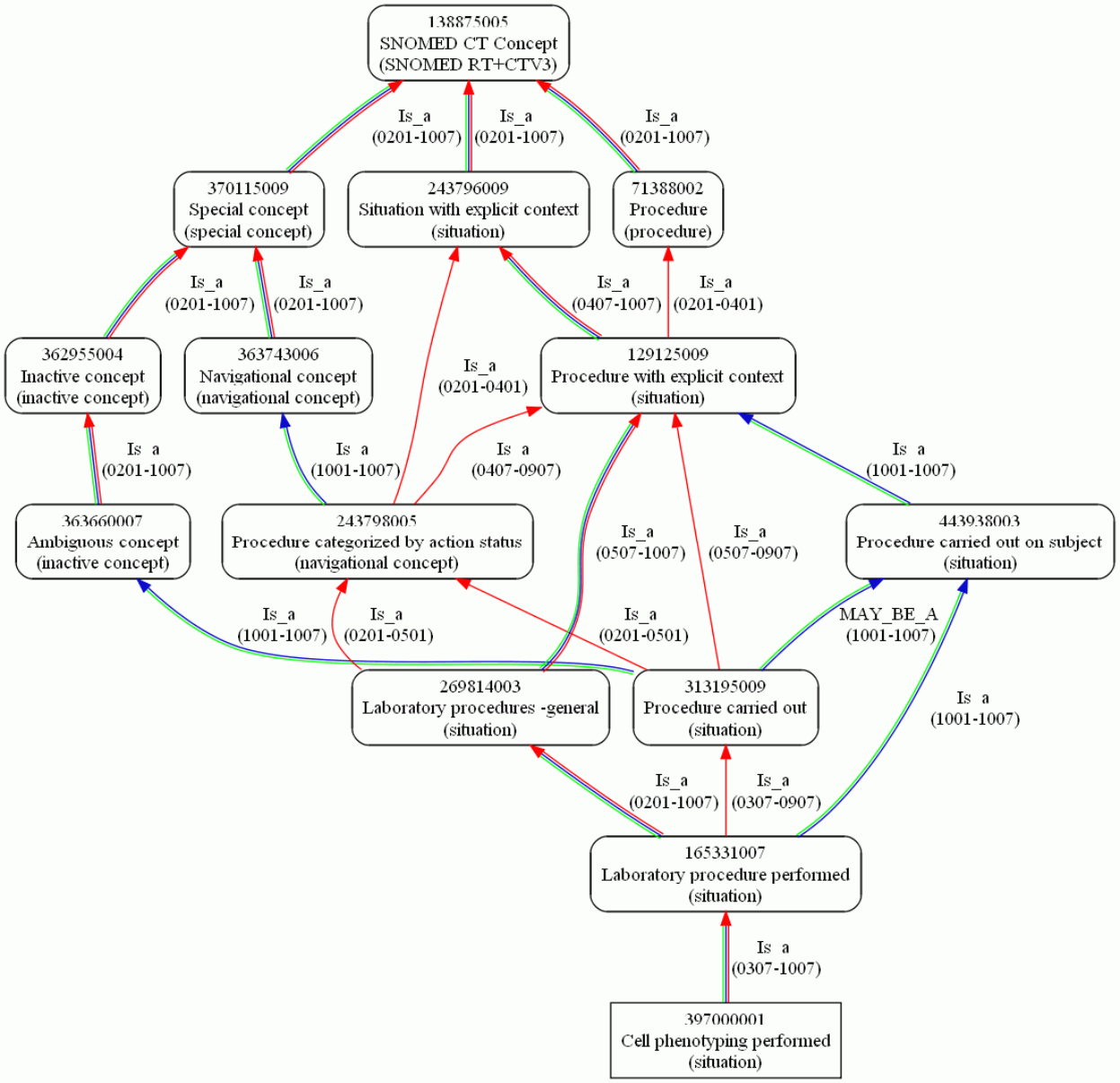

Figure 1. Impact of SNOMED CT revisions on the classification of SNOMED CT concept 397000001 with Fully Specified Name '*Cell phenotyping performed (situation)*'.

Methods

The study presented here focused on a subset of 883 SNOMED CT concepts – henceforth referred to as *source concepts* – used within a cancer clinic for encoding synoptic pathology reports and tumor registry data and for querying a biospecimen repository, all together covering almost 16,000 occurrences related to 10,000 unique patients.

Table 1. Transitive closure sets for the source concept '44228008: *Surgical margins involved by tumor* (finding)'

Rel-ID	Version																		Rel-Type	Target Concept
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
H-18608	6	6	7	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Is a	SNOMED CT Concept (SNOMED RT+CTV3)
H-23694	5	5	6																Is a	Finding (finding)
H-18607	4	4	5																Is a	Finding by method (finding)
H-12792	3	3	4																Is a	Test finding (navigational concept)
H-12789	3	3	3																Is a	Laboratory test finding (navigational concept)
H-07371	2	2	2																Is a	Sample finding (finding)
H-07373	2	2																	Is a	Morphologic finding (finding)
220039029	1	1																	Is a	Clinical sample finding (finding)
H-07370			3																Is a	Histopathology finding (finding)
H-07368			2																Is a	General pathology (finding)
H-07369			2																Is a	Laboratory finding present (navigational concept)
2030386023			1																Is a	Pathology examination findings present (finding)
2030387025			1																Is a	Surgical margin finding (finding)
H-18182				8	8	8	8	8	8	8	8	8	8	8	7	6	6	6	IsA	SNOMED CT Concept (SNOMED RT+CTV3)
H-27662				7															IsA	Finding (finding)
H-23376				6	6	6	6	6	6	6	6	6	6	6					IsA	Finding by method (finding)
H-18183				5	5	5	5	5	5	5	5	5	5	5	5				IsA	Test finding (navigational concept)
H-07379				4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	IsA	Histopathology finding (finding)
H-12412				4	4	4	4	4	4	4	4	4	4	4	4				IsA	Laboratory test finding (navigational concept)
H-07380				4	4	4	4	4	4	4	4	4	4	4	4				IsA	Sample finding (finding)
H-12418				4	4	4	4	4	4	4	4	4	4	4	4				IsA	Morphologic finding (finding)
H-07378				3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	IsA	General pathology (finding)
H-12793				3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Is a	Special concept (special concept)
H-07377				3	3	3	3	3	3	3	3	3	3	3	3				IsA	Clinical sample finding (finding)
H-07381				3	3	3	3	3	3	3	3	3	3	3	3				IsA	Laboratory finding present (navigational concept)
H-07374				2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Is a	Inactive concept (inactive concept)
H-07382				2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	IsA	Pathology examination findings present (finding)
H-07384				2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	IsA	Surgical margin finding (finding)
2228147020				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Is a	Duplicate concept (inactive concept)
2295897028				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	SAME AS	Surgical margin involved by tumor (finding)
H-18186					7	7	7	7	7	7	7	7	7	7	6	5	5	5	IsA	Clinical finding (finding)
H-12417																4	4	4	IsA	Evaluation finding (finding)

Legend. ***Rel-ID***: relationship ID, either an original component ID from SNOMED CT or generated during computation (the latter preceded by 'H-'); ***Version***: digits represent the minimal path length, blank when the relationship is not present in a version; ***Rel-Type***: either an original relationship type from SNOMED CT (here 'Is a' when path length equals '1' and 'SAME AS') or a computed one following the transitivity principles outlined in Table 2; ***Target Concept***: the Fully Specified Name of a SNOMED-CT concept in the transitive closure set of the source concept.

For each source concept, all concepts – from here on referred to as *target concepts* – within the transitive closure set of the *Is a* relation and all historical relations – *Was A*, *Replaced By*, *Same As*, *May Be*, *Moved To*, and *Moved From* – were computed for each SNOMED CT version from January 2002 to July 2010, together with their concept status and path length to the source concept. Computing the transitive closure set involved traversing the target of each of these relationships included in SNOMED CT's Relationships Table of each version to look for and follow further relationships until all paths through the hierarchy reach the root concept (closure). When a target concept could be reached by traversing more than one path, the shortest path length from source concept to target concept was preserved. Table 1 shows these computations for the source concept '44228008: *Surgical margins involved by tumor* (finding)'. Table 2 displays the rules used to compute the composite relationships during the transitive closure computation of this concept.

In a second step – again for each version of SNOMED CT – the *genericity* of each target concept was computed, where *genericity* was defined as the number of times a target concept appears in the paths from all source concepts to the root of the vocabulary. The maximum possible value for genericity, under this definition, was 883, i.e. the number of source concepts. These values were then used to compute, for each source concept (SC) its *information content* defined as the sum of the values obtained by dividing the genericity of each target concept TC on a path from SC to the top by the respective path lengths from SC to TC. Table 3, as an example, shows the results for the concept '*pN1b: Metastasis in internal mammary lymph nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent (breast) (finding)*'.

Table 2. Transitivity rules for relationships with distinct signature.

C1→C2	C2→C3	C1→C3
Is a	Is a	Is a
Is a	IsA	IsA
IsA	Is a	IsA
IsA	IsA	IsA
SAME AS	Is a	IsA
SAME AS	IsA	IsA
Is a	SAME AS	IsA
IsA	SAME AS	IsA

Table 3. Example of the calculation of the information content of a source concept.

Target Concepts	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18
SNOMED CT Concept (SNOMED RT+CTV3)	0	58	56	103	139	104	104	104	93	93	93	93	93	93	93	93	93	93
Staging and scales (staging scale)	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tumor staging (tumor staging)	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cancer staging (tumor staging)	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tumor-node-metastasis (TNM) tumor staging system (tumor staging)	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Generic tumor staging descriptor (tumor staging)	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tumor-node-metastasis (TNM) classification of malignant tumor after operation (observable entity)	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N category (observable entity)	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pN1 category (finding)	0	3	5	10	11	11	11	11	11	11	11	11	11	11	11	11	12	12
pN1: Metastasis in 1 to 3 axillary lymph nodes, and/or in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent (breast) (finding)	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Finding (finding)	0	0	19	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Clinical history and observation findings (finding)	0	0	16	54	0	54	54	54	46	46	46	46	46	46	46	46	46	46
Clinical finding (finding)	0	0	19	64	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tumor finding (finding)	0	0	22	80	81	81	81	81	65	65	65	65	65	65	65	65	65	65
Node category finding (finding)	0	0	5	10	11	11	11	11	11	11	11	11	11	11	11	11	11	11
Tumor stage finding (finding)	0	0	18	0	0	0	0	0	56	56	56	56	56	56	56	56	56	56
Tumor-node-metastasis (TNM) tumor staging finding (finding)	0	0	22	71	73	73	73	73	73	73	73	73	73	73	73	73	73	73
pN category finding (finding)	0	0	6	14	15	15	15	15	15	15	15	15	15	15	15	15	15	15
N1 category (finding)	0	0	3	7	8	8	8	0	0	0	0	0	0	0	0	0	0	0
Breast TNM finding (finding)	0	0	10	10	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Clinical finding (finding)	0	0	0	0	88	63	63	63	55	55	55	55	55	55	55	55	55	55
Finding of lesion (finding)	0	0	0	0	0	65	65	65	54	54	54	54	54	54	54	54	54	54
pN1b category (finding)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4
Total Information content	0	123	203	484	440	500	500	492	493	493	493	493	493	493	493	493	498	498

The sum of the information contents of all source concepts within a version would then yield the information content of that entire version. *Our first hypothesis is that the evolution of the information content of the versions over time can be used as an indicator to decide whether to upgrade to a new version.*

Intermediate inspection of these results suggested that the procedure thus far described could be used to detect possible mistakes in SNOMED CT. The grey shaded cells in Table 3 do indeed show that in some versions target concepts for the source concept disappeared from the transitive closure set while reappearing in later versions. It was also discovered that when target concepts permanently disappeared from the transitive closure set, this could not always be explained by the retirement of the target concept within the corresponding version. Although this does not mean that there is a mistake – it might rather be the correction of a mistake – it was decided to register this and similar phenomena as a *suspicious event*. Each source concept / target concept pair was therefore additionally marked as being the seat (or not) of such an event and for each version tallies for such events were computed for all such events over all previous versions until another change was effected. Thus if a change was marked in some version as being a suspicious event, it stayed marked as such until in some later version – if at all – another change occurred that no longer met the requirements for being suspicious (Table 4). An extreme example is the concept '*64766004*: *Ulcerative colitis (disorder)*' which saw its *isa*-relation with '*24526004*: *Inflammatory bowel disease (disorder)*' change five times. *Our second hypothesis is that evolution of these tallies over time, the suspicious event perseverance, yields a second indicator for migrating to a new version of SNOMED CT.*

Table 4. Evolution of suspicious events

	N	Binary %
Unmarked	15689	
Stay unmarked	11182	71.27%
Become suspicious	4507	28.73%
Stay suspicious	1812	40.20%
Become unmarked	2695	59.80%
Stay unmarked	2296	85.19%
Become suspicious	399	14.81%
Stay suspicious	332	83.21%
Become unmarked	67	16.79%
Stay unmarked	66	98.51%
Become suspicious	1	1.49%
Stay suspicious	0	0.00%
Become unmarked	1	100.00%

A last step of the analysis was to compare these results with the quality metrics obtained by applying *Evolutionary Terminology Auditing* to the source and target concepts in the transitive closure set. This method uses a more recent version of a terminology or ontology as gold standard to compare with earlier versions [14]. Differences are expressed in terms of four major categories – justified presence (JP), justified absence (JA), unjustified presence (UP) and unjustified absence (UA) – which further give rise to seventeen subcategories reflecting all possible combinations determined by whether changes in a new version are motivated by changes (1) in reality, (1) in the ontology authors' understanding of reality, or (3) by editorial mistakes. With each of these subcategories corresponds an error value (e_i) with a minimum magnitude of 0 (no mistake, which is the case for all subcategories of JP and JA) and maximum of 5 [14]. The overall quality of an earlier version with respect to a chosen later version is then computed by means of formula (1):

$$\frac{\sum_{i=1}^{JP+UP}(5-e_i)}{5(JP+UP)+4UA} \quad (1)$$

Results

The 883 source concepts studied were linked by means of 15,689 relationships to 1,415 target concepts which is only a small fraction of the total number of concepts in SNOMED CT. Of the 15,689 relationships, 28.73% were found to be suspicious (Table 4).

Figure 2 shows the evolution of the information content of the source concepts over time. The biggest increase in information content occurs over the first few versions, with the exception of version v5 (January 2004). Also version v17 (January 2010) shows a minor (though not noticeable on the chart) increase: from 384,960.7605 to 385,449.781. Figure 3 illustrates the suspicious event perseverance of the source concepts over time. Significant changes in the suspicious event perseverance are those which constitute a downwards trend break, thus a reduction in the perseverance. This is here the case for versions v6 (July 2004), v7 (January 2005), and v14 (July 2008).

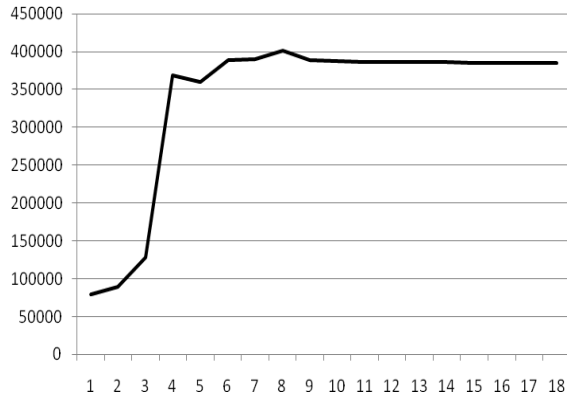

Figure 2. Evolution of the information content of all source concepts over all versions.

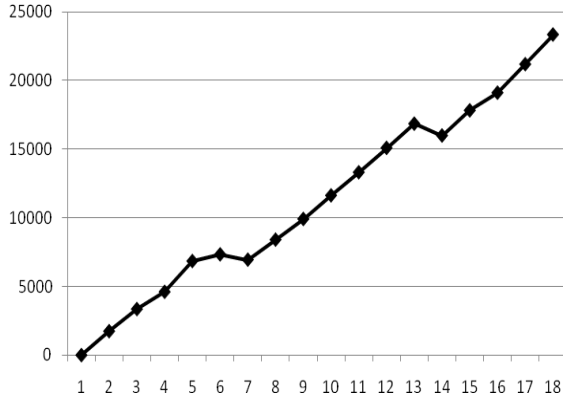

Figure 3. Evolution of suspicious event perseverance of all source concept/target concept.

Both indicators together thus suggest that applications using the set of source concepts studied do not benefit from upgrades to SNOMED versions issued between July 2005 and January 2008 (v8 – v13), nor from both 2009 versions (v15 – v16), nor from the July 2010 version (v18).

Figure 4 displays two distinct views on the evolution of the quality of the transitive closure set computed by means of formula (1).

As argued for and motivated in [14], unjustified absences all have an error magnitude of 1. All unjustified presences in the transitive closure set being the result of unintentional encoding rather than misunderstanding what is the case in reality, the applicable error magnitude is for these cases also 1.

The Qlv-view uses the most recent version (July 2010) as gold standard; thus when the difference between two consecutive points on this curve shows a large increase, as is for instance the case between v3 and v4, then this means that, from the perspective of v18, it would have been worthwhile to upgrade from v3 to v4. The curve shows a gradual increase of the quality primarily in the first four versions and to a lesser extent in the next four.

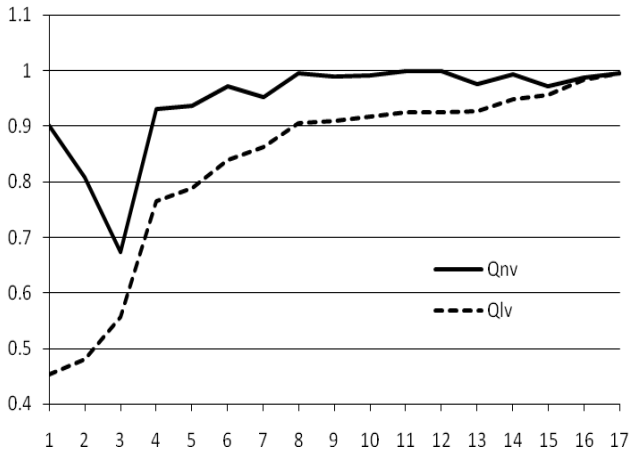

Figure 4. Quality evolution of concepts in the transitive closure set using the next version (Qnv) or last version (Qlv) as gold standard.

The Qnv-view, in contrast, uses as reference the version which follows immediately the version for which the quality is computed. Thus when the difference between two consecutive points on this curve shows a large increase, as is the case between v3 and v4, then this means that it is worth upgrading to the *next* version, thus v5. The Qnv-view suggests – roughly – that v5 would certainly, and v7 to a lesser extent, justify an upgrade and that, starting with v8, the quality oscillates around 97.5%, thus not motivating any further upgrade.

Note that the information to generate Qlv – in contrast to Qnv, the information content and suspicious event perseveration – is only available at the time when v18 is released and is thus not useful as a decision aid prior to that time. Furthermore, as witnessed by the oscillating behavior of Qnv, the assumption that the most recent version is the best gold standard does not seem to hold completely.

Table 5, finally, offers a global overview of the sorts of strategies that can be used to make a decision on whether to upgrade to a new version of SNOMED CT once available.

Table 5: Various strategies to decide on upgrading and comparison with '*last version quality improvement*'.

Version		Information content	Suspicious event perseverance	*'Next version'* quality improvement	*'Last version'* quality improvement					
					LQV	5%	10%	15%	20%	25%
v1	January 2002	79,424.5499	↗	0.902158875	0.45334					
v2	July 2002	88,838.2435	↗	0.807276754	0.47980	Y				
v3	January 2003	127,864.0057	↗	0.674182895	0.55653	Y	Y	Y	Y	
v4	July 2003	368,261.2260	↗	**0.930927215**	0.76591	Y	Y	Y	Y	Y
v5	January 2004	360,409.8979	↗	**0.936412939**	0.78907					
v6	July 2004	388,775.7097	↘	**0.970773297**	0.83995	Y				
v7	January 2005	390,637.7034	↘	0.952138559	0.86259		Y			
v8	July 2005	401,196.6372	↗	**0.995326645**	0.90446	Y		Y		
v9	January 2006	389,021.8138	↗	0.988407635	0.90854					
v10	July 2006	387,226.2485	↗	0.992040473	0.91776					
v11	January 2007	386,714.4974	↗	**0.999406567**	0.92518				Y	
v12	July 2007	386,816.9581	↗	0.998631853	0.92567					
v13	January 2008	386,803.0803	↗	0.976249334	0.92638					
v14	July 2008	386,075.3172	↘	0.992155217	0.94874					
v15	January 2009	384,952.4653	↗	0.971849459	0.95583	Y	Y			
v16	July 2009	384,960.7605	↗	0.987649670	0.98262					Y
v17	January 2010	385,449.7811	↗	0.994921958	0.99492					
v18	July 2010	385,052.7645	↗	-	-					

The first strategy (column 3 of Table 5, displayed values representing for each version the computed information content of the subset) is to use absolute increase in information content: grey shaded values indicate a version for which the absolute information content is greater than the one computed for the previous version.

The second strategy is to upgrade when there is a downward trend break in the suspicious event perseverance.

The third strategy is to upgrade when the magnitude of the quality of a new version with respect to the previous version according to formula (1) is higher than any corresponding magnitude obtained earlier from comparing previous consecutive version pairs. Whenever a value satisfies this requirement, it is printed in bold. As explained above, it is in such case the *next* version which it is worthwhile upgrading to. These versions, too, are highlighted with a grey background.

Further strategies can easily be derived by combining any of the three just described.

Table 5 offers at the same time a way to compare whether a strategy has proven to be successful in light of how much better in quality the very last version is compared to any of its predecessors – these quality values appear in the column labeled 'QLV' – and under the further constraint that for cost-benefit reasons an upgrade would only be considered if a quality increase is obtained of at least 5%, 10%, and so forth. Whenever such a condition is met, a 'Y' is registered in the cell corresponding to the qualified version and the desired cost-benefit. As an example, the v5 cell of the 5% cost-benefit column is *not* marked 'Y,' because the increase in QLV compared to v4, i.e. the most recent version with respect to v5 for which the requirement was met, is less than 5%. In contrast, v6 *is* so marked, because this is the first version after v4 for which the requirement *is* met.

As can be seen, the information content strategy approximates closely the 5% quality increase requirement, the three deviations being (1) unnecessary upgrades in January 2005 and July 2007, and (2) a failure to upgrade in January 2009 which is corrected in January 2010. Combining this strategy with the suspicious event perseverance strategy would have led to an upgrade in July 2008 instead of January 2009.

Discussion

The term 'information content' is used in ontology contexts with a number of distinct meanings. Most often it designates a metric which captures how often a representational unit or variants thereof appears in *external* corpora [15]. Of a different nature are information content metrics which rely solely on the *internal* structure of an ontology and which are based on the idea that the higher a representational unit appears in the taxonomy, the lower its

information content is [16]. The metric proposed here belongs to the latter sort, but rather than being used for semantic similarity computations [17] it elaborates on the idea, advanced in [18], that it might be used for quality control in ontology development by following the degree to which the information content of individual concepts changes over time. Hogan and Slee used *Shannon's information entropy* which is somehow related to information content, to suggest the use of SNOMED CT instead of ICD-9-CM for coding diagnoses [19]. Measures for information content have thus far not been used to assess whether it is worthwhile to upgrade from one version of an ontology to another.

An increase in information content from one version to another as defined here can be brought about by several sorts of changes. Within the context of SNOMED CT and the subset thereof studied here, such changes can be the introduction of new representational units and the creation or elimination of relationships between intermediary representational units along the transitive closure paths of the source concepts. It is the increase in the number of relationships which in the set of source concepts studied here is responsible for the substantial information content increase in the first seven versions. Situations like this will occur in the first place when data repositories make use of codes drawn from classification systems that at some point in time become integrated in SNOMED CT but for which it takes several versions before all relationships with other SNOMED CT concepts are added.

The notion of *suspicious event perseverance* as defined here is new. It functions as an indicator warning of the appearance of representational units and of assertions about relations that obtain between the referents of these units – in SNOMED CT this means *concepts* and *relationships* – or, more importantly, the disappearance or reappearance thereof in later versions. Such changes do not always indicate mistakes of some sort: as discussed in *realism-based ontology evolution* [20], *realism-based ontology matching* [21] and *realism-based terminology auditing* [14], no mistakes are committed when such changes in ontologies mimic changes in the corresponding referents. When that is not the case, however, the appearance of a representational unit or relational assertion might either involve the correction of an unjustified absence or the introduction of an unjustified presence, whereas the disappearance might either involve the correction of an unjustified presence or the introduction of an unjustified absence. To determine which is the case requires careful manual inspection. This uncertainty might, erroneously, be perceived as a limitation of our methodology but, as we have demonstrated here, this manual inspection is not necessary to make decisions on whether or not to upgrade: it is the evolution of the suspicious events rather than the events themselves that has predictive power and this evolution can be computed automatically.

A limitation of our method, perhaps, is the computational power required for these calculations. Transitive closure computations place heavy demands on computer resources specifically for very large sets of source concepts. New computational techniques such as incremental reclassification over versions will however likely make this limitation go away [22].

The predictive power of the strategy which combines information content increase and suspicious event perseveration is striking. Of course, this conclusion is based on only one use case. Further tests are required to assess whether the same results will be obtained with other subsets, for instance, the subset of SNOMED CT codes organizations are required to use to satisfy the current Meaningful Use criteria.

A reasonable question to ask is whether there is practical value in applying this methodology when Meaningful Use criteria might evolve in such a way that organizations will be forced to upgrade to new versions of SNOMED CT immediately when these are introduced. Here we argue that our results indicate that imposing such obligation blindly might be a foolish idea leading to unnecessary effort on behalf of healthcare providers. A better approach would be that decision makers in these matters would use the methodology here proposed to test with each new release whether upgrading would make sense.

Finally, the notion of suspicious event perseverance might not only be useful for the purpose of decision making with respect to the application of a new version, but also for quality assurance purposes concerning the further development of SNOMED CT itself. Specifically, the temporal disappearance of a target concept from a transitive closure set raises questions about the adequacy of the internal quality assurance principles and classification engine used in the SNOMED CT authoring environment. We found 500 occurrences of such temporal disappearances, and in light of the 15,689 individual relationships this might seem like only a small fraction. In fact, however, it is quite large as these erroneous deletions originate from only 833 source concepts, and are targeted towards only 1415 target concepts. These findings are consistent with other recent quality evaluations of SNOMED CT that make use of description logics [23].

Conclusion

Two indicators, the evolution of the global information content of an ontology over consecutive versions, and the perseverance of what were called *suspicious events*, are proposed to assess whether it is worthwhile upgrading when a new version of SNOMED CT is released. The indicators can be computed fully automatically and are statistically unrelated. Trend breaks in the evolution of the indicators are suggestive for the possible benefit of an upgrade. Comparison with a realism-based quality metric demonstrates that this methodology is successful when applied to the subset of SNOMED CT codes used for coding pathology reports. It is argued that the methodology is also useful for internal quality control of SNOMED CT.

Acknowledgements

The work described was funded in part by grant R21LM009824 from the National Library of Medicine. The content of this paper is solely the responsibility of the author and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

References

1. Donnelly K. SNOMED CT: The advanced terminology and coding system for ehealth. In: Bos L, Roa L, Yogesan K, O'Connell B, Marsh A, Blobel B, editors. Studies in Health Technology and Informatics - Medical and Care Compunetics 3 Vol 121. Amsterdam: IOS Press; 2006. p. 279 - 90.
2. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of Information in Medicine. 1998;37(4-5):394-403.
3. Sampalli T, Shepherd M, Duffy J, Fox R. An evaluation of SNOMED CT® in the domain of complex chronic conditions. International Journal of Integrated Care. 2010 March 24;10:e038.
4. Park H, Lundberg C, Coenen A, Konicek D. Evaluation of the content coverage of SNOMED-CT to represent ICNP Version 1 catalogues. Stud Health Technol Inform. 2009;146:303-7.
5. J.E. Andrews, R.L. Richesson, J. Krischer. Variation of SNOMED CT coding of clinical research concepts among coding experts. Journal of the American Medical Informatics Association. 2007;14(4):497-506.
6. M.F. Chiang, J.C. Hwang, A.C. Yu, D.S. Casper, J.J. Cimino, J. Starren. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. AMIA 2006 Symposium Proceedings2006. p. 131-5.
7. Schulz S, Cornet R. SNOMED CT's Ontological commitment. In: Smith B, editor. ICBO: International Conference on Biomedical Ontology. Buffalo NY: National Center for Ontological Research; 2009. p. 55-8.
8. Ceusters W, Smith B, Kumar A, Dhaen C. Ontology-based error detection in SNOMED-CT®. In: Fieschi M, Coiera E, Li Y-CJ, editors. MEDINFO 2004. Amsterdam, The Netherlands: IOS Press; 2004. p. 482-6.
9. Wei D, Halper M, Elhanan G, Chen Y, Perl Y, Geller J, et al. Auditing SNOMED relationships using a converse abstraction network. AMIA Annu Symp Proc 20092009. p. 685-9.
10. Hartung M, Kirsten T, Rahm E. Analyzing the evolution of life science ontologies and mappings. Leipzig: Interdisciplinary Centre for Bioinformatics, 2008 Contract No.: 17.
11. Ceusters W. Applying Evolutionary Terminology Auditing to SNOMED CT. American Medical Informatics Association 2010 Annual Symposium (AMIA 2010) Proceedings. Washington DC, 2010. p. 96-100.
12. Wade G, Rosenbloom T. The impact of SNOMED CT revisions on a mapped interface terminology: Terminology development and implementation issues. Journal of Biomedical Informatics. 2009;42(3):490–3.
13. Elhanan G, Perl Y, Geller J. A survey of direct users and uses of SNOMED CT: 2010 Status. AMIA Annu Symp Proc2010 p. 207-11.
14. Ceusters W. Applying evolutionary terminology auditing to the Gene Ontology. Journal of Biomedical Informatics; Special Issue on Auditing of Terminologies. 2009;42(3):518-29.
15. Resnik P. Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. Journal of artificial intelligence. 1999;11:95-130.
16. Seco N, Veale T, Hayes J. An intrinsic information content metric for semantic similarity in WordNet. Proceedings of ECAIf2004, the 16th European Conference on Artificial Intelligence, 2004. p. 1089-90.
17. Sánchez D, Batet M, Iserna D. Ontology-based information content computation. Knowledge-Based Systems. 2011;24(2):297-303.

18. Van Buggenhout C, Ceusters W. A novel view on information content of concepts in a large ontology and a view on the structure and the quality of the ontology. International Journal of Medical Informatics. 2005;74(2-4):125-32.
19. Hogan W, Slee V. Measuring the information gain of diagnosis vs. diagnosis category coding. AMIA Annu Symp Proc 2010 2010. p. 306-10.
20. Ceusters W, Smith B. A realism-based approach to the evolution of biomedical ontologies. Biomedical and Health Informatics: Proceedings of the 2006 AMIA Annual Symposium. Washington DC: American Medical Informatics Association; 2006. p. 121-5.
21. Ceusters W. Towards A realism-based metric for quality assurance in ontology matching. In: Bennett B, Fellbaum C, editors. Formal Ontology in Information Systems. Amsterdam: IOS Press; 2006. p. 321-32.
22. Grau BC, Halaschek-Wiener C, Kazakov Y, Suntisrivaraporn B. Incremental classification of description logics ontologies. Journal of Automated Reasoning. 2010 April 2010;44(4):337-69.
23. Rector A, Brandt S, Schneider T. Getting the foot out of the pelvis: modelling problems affecting use of SNOMED-CT hierarchies in practical applications. Journal of the Amercian Medical Informatics Association. 2011;18(4):432-40.