

Response to AMIA 2010 Reviewers¹

AMIA-1358-A2009 - Paper
Applying Evolutionary Terminology Auditing to SNOMED CT
Ceusters W

Dear Prof. Ceusters:

This letter reports the decision regarding your paper submitted to the AMIA 2010 Annual Symposium, as well as the process for revision of accepted papers. Submissions for the Annual Symposium remain constant on a yearly basis.

AMIA 2010 Decision: Accept/Proceedings Eligible

There is an additional step in the paper preparation process if that will help ensure the scientific validity of the papers and will confirm the status of AMIA's papers as fully peer-reviewed. Authors whose papers have been accepted for presentation at the meeting are asked to address carefully the comments of the reviewers and the SPC, and to submit a revised version of their manuscript for publication in the Proceedings. We are not asking authors necessarily to accept all suggestions; but when submitting their revised manuscripts, authors can explain why they did and did not respond to specific comments. Space will be provided for this explanation within the final submission area the website. The SPC reserves the right to review authors' revisions and the reasoning behind them, and may remove from the program papers of authors whose changes are not responsive to the comments or who otherwise do not comply with this final submission procedure.

[...]

Thank you again for your contribution to the AMIA meeting. AMIA will share the date, time, and other logistical information specific to your AMIA 2010 presentation in late July. Please contact Jeff Williamson by e-mail at jeff@amia.org if you have any questions about any aspect of this process.

Sincerely,

Gil Kuperman, MD, PhD
Chair, AMIA 2010 Scientific Program Committee

Reviewer: 1

Comments to the Author

The author previously proposed the Evolutionary Terminology Auditing (ETA) approach to measure the changes among different versions of biomedical terminologies. ETA was

¹ Reviewers' and editor's comments are printed in blue, my responses in black

used to check the Gene Ontology, and in this paper the author applied ETA to quantify the changes in versions of SNOMED CT. The paper is clearly written, and results seem reliable as well as useful for the SNOMED CT's further development.

The problem I have lies in the ETA model itself, rather than the specific work done in this paper. What ETA captures are basically two types of changes in different versions of an ontology, changes in reality, and changes of the ontology modelers' knowledge about the reality.

It seems to me that a very important type of changes is missing in ETA, which comes from the changing modeling choices, e.g., the modelers may want to make a more detailed classification for some concept in the new version, thus introducing new concepts and isa relations and removal some concepts or isa relations. Such changes account for a notable proportion of ontology evolving, especially when the modeling of some biomedical entities (for example, mass, count, and fuzziness) is still open problem currently.

→ But this is exactly what is covered in this paper, nothing additional. If an author wants to *'make a more detailed classification by introducing new concepts'*, then this is for sure because either reality, or his understanding of this reality including issues of relevance changed. Specifically for entities which pose a *'still open problem'*, it is the understanding that changes.

The ETA assumes that all changes fall into the categories as described in Table 2,

→ that is correct. The exception would be changes in the ontology because of changes in the representation ('modeling') formalism. That did not happen in SNOMED CT. I made this clear in the paper, by adding the sentence *'Note that changes in the RUs because of changes in the representation formalism itself do not count as what we mean by 'changes' in this paper.'*

thus the absent of modeling choice changes in ETA makes the statistic results doubtful.

→ I disagree for reasons just stated.

Reviewer: 2

Comments to the Author

The author describes an interesting methodology to assess "how much is a new version of a terminology (SNOMED CT) better than any previous version". The paper raises awareness about the need for better documentation of the reasons for the changes made to the terminology, and the continuity of the line of work described in the submission might contribute promising insights to objectively assess the evolution of the terminology after the methodology is validated and its potential contribution to other evaluation methods established.

However, the Evolutionary Terminology Auditing technique faces significant challenges as a structural metric to assess SNOMED CT quality changes over time. While the rationale underlying the proposed framework seems reasonably solid, it is unclear whether many of the assumptions made in the paper are valid.

→ of course they are not. I made that more explicit.

For example, SNOMED CT editorial direction, priority setting and allocation of authoring resources are the result of a consensus-based process involving many member countries. The number of quality improvement initiatives largely exceed the available resources, and therefore changes between release do not necessarily reflect the “author’s understanding or the state of the art.

→ I agree, in general. I added a sentence to that effect: *‘For sure, the assumptions described in the methods section are not valid from one version to another and the statistics obtained need to be assessed in that light. Lack of resources might for instance prevent changes to be introduced although the authors know it has to be done at some point.’*

For example, more than 20,000 limited concepts were included in the initial release to support legacy applications (rather than resolving unjustified absences of content), and most of them were retired in January 2010 after support was no longer needed.

→ so there is a change in the relevance, that is covered in the paper

The merge between Clinical Terms V3 and SNOMED RT caused the retirement of a massive amount of content to enable better traceability between the history of those terminologies and SNOMED CT.

→ that is again a relevancy issue and is thus covered by our method.

The author cites a very good example related to the management of anomalies produced by the inappropriate merging of “synonyms” in the original terminologies, where concepts that were not intrinsically ambiguous were retired as such, because that would minimize the impact on existing implementations (rather than reflect the state of the art on what ambiguity is).

It is unclear whether the addition of new content reflects unjustified absence, since a significant part of the requests for new content are associated with excessive pre-coordination patterns,

→ I do not understand what the reviewer has in mind here. If he means that because of an existing pre-coordinated concept in v1 implicitly denoting ‘A isa C which B’, B is added in v2, then I believe that B was then unjustifiably absent in v1.

while there are significant areas with known issues that were not remodeled because of lack of resources or other priorities

→ and thus less relevant ? (what is covered)

(so “unchanged content” is not necessarily a marker of good quality).

→ I agree with that.

When the author states that “we rely on the assumption that with each release its authors assume ... that all its constituent expressions are of the correct type”, the assumption might not be a valid one.

→ I agree and made it clear.

Some of the numbers in table 1 would require a revision. For example, the number of concepts retired without a specified reason seems too low.

→ I checked, but it is correct over the period studied.

Sadly, one of the main reasons for change (changes in editorial policy (the proxy for “author’s understanding”)) is not reflected in SNOMED history tracking (for example the remodeling of situations with explicit context, the reorganization of some hierarchies, etc.). As the author suggests, the lack of better documentation of the reason for change severely limits external analysis and contributions, and particularly, it limits the validity of the mapping between the status changes and the reality-based interpretation.

→ let us hope they do something about it ☺

Finally, the quality of a terminology resource like SNOMED CT can also be interpreted on how fit for a specific purpose it is. For example, granularity appropriate for general medicine might not be adequate for specialized care or research. It is very likely that the assessment of quality would require multiple dimensions rather than an one-dimensional assessment.

→ that is included in the methodology: note that relevancy-to-purpose is one of the three parameters used in our metric (see table 2)

As the author discusses, a better insight in the concrete reasons for change would have given a more accurate application of the proposed metric. I encourage the author to seek a better understanding or propose improvements about the required changes to the IHTSDO request submission process, as that would help remove confounding factors originated in the subject of analysis and would help to assess the real value of the proposed technique.

→ it would, but as for the resources available to SNOMED CT authors, mine are limited too.

Reviewer: 3

Comments to the Author

Nice paper contributing to the quantification of quality aspects of terminologies.

It would be nice if some light could be shed on the parts of the hierarchy in SNOMED CT with the majority of the changes. Taking SNOMED CT as a whole in my opinion leads to an underestimate of changes. There will very likely be hierarchies with relatively much change (findings, procedures) and with little change (qualifiers, anatomy). It would be good to address this, even though I realize this can't be worked out in full in this conference paper.

→ we agree with the reviewer. That is work in progress.

Textual comments:

In the discussion you write: "The figures seem to indicate that with respect to concepts, only small quality improvements are introduced with each new version, i.e. roughly 2% with an overall quality improvement of about 16%"

Then in the conclusion you write: "This need not be a negative finding for two reasons: (1) the proposed metric becomes less sensitive when the size of the terminology increases, and (2) it might very well be that SNOMED CT 'got it right' from the very beginning, since, after all, its real foundations were created almost 50 years ago."

This sentence should be moved from the conclusion (which it isn't) to the discussion, right after the paragraph copied above.

→ we did so.

In the discussion:

"The move of SNOMED CT to migrate brand-named products to subsets eliminates this problem."

Should be "... to reference sets ..." or "... to extensions ..." instead of "... to subsets ..."

→ has been corrected

"Our mapping is this the best estimate that we could make"

Correct this sentence

→ we did so.

In the conclusion:

(2) to what degree reflect terminology changes evolutions in the underlying domain or the terminology authors' understanding thereof.

Rewrite as "(2) to what degree do terminology changes reflect evolutions ..."

→ we did so.

Generally, ensure that the conclusion concludes. Move the discussions to the discussion section.

→ we did so.

References:

Rather than having 50% of references to work from your own group, it would be good to include references to the work of others on evaluating changes in SNOMED, e.g.,

G. Wade and T. Rosenbloom, The impact of SNOMED CT revisions on a mapped interface terminology: terminology development and implementation issues, J Biomed Inform 42 (3) (2009), pp. 490–493.

→ we know this paper, but it is not relevant for our purposes here.

Track Chair: 1

Comments to the Author:

(There are no comments)

Review Chair: 2

Comments to the Author:

(There are no comments)