

Creating a SNOMED CT® Genetic Disorder Subset

John S. CARTER^{a,1}, Manasi M. NABAR^a
^a*Apelon, Inc., Ridgefield, CT, USA*

Abstract. The use of standardized terminologies to capture and share clinical data can facilitate patient care and decision support, required reporting, and analytic use cases. We report on the creation of a subset of genetic disorder concepts from SNOMED CT, based on a hybrid lexical and semantic query and the correspondence between its results and the identification of a subset derived from mappings to the Genetics Home Reference website. We conclude that SNOMED CT can be used to define a genetic disorder subset, but that coverage gaps and granularity mismatches with non-SNOMED-based resources still exist.

Keywords. SNOMED CT, vocabulary, ontology, terminology, genetic disorders

1. Motivation

The use of standardized terminologies to capture and share clinical data can facilitate patient care and decision support, required reporting, and analytic use cases. SNOMED CT®, produced by the International Health Terminology Standards Development Organisation (IHTSDO, <http://www.ihtsdo.org>) is emerging as the standard for a wide range of data capture and reporting use cases around the world.

On January 28th, 2011, IHTSDO announced a new “public good” use policy,[1] which allows free worldwide use of SNOMED CT identifiers and descriptions when that usage is considered to be in the public good. The announcement referred to the standardization on SNOMED CT disorders within international genetic databases produced by the National Center for Biotechnology Information (NCBI) as an early benefit of the new policy. Motivated by the IHTSDO public good usage policy announcement, we set out to define and evaluate a method for creating a subset of SNOMED CT concepts to identify genetic disorders.

2. Results

We used the July 2010 version of SNOMED CT in English and the open source Distributed Terminology System (available: <http://apelon-dts.sourceforge.net>) software to define a subset of genetic disorders concepts. While SNOMED CT includes a rich hierarchical organization for disease concepts, there is no single hierarchy branch that covered the gamut of genetic disorders. There are partially overlapping but not

¹ Corresponding author: John S. Carter, MBA, Vice President, Apelon, Inc., 100 Danbury Rd. Ste. 202, Ridgefield, CT, 06877, USA; E-mail: jcarter@apelon.com

comprehensive categories for “Congenital Disease”, “Familial Disease” and “Hereditary Disease”.

Therefore, as shown in Figure 1, we used the DTS Subset Expression Editor to query SNOMED CT using both lexical (e.g., “HAVING CONCEPT NAME OR SYNONYM MATCHING *familial*”) and semantic (e.g., “ALL descendants of Hereditary disease (disorder)”) features of SNOMED CT. This query generated a candidate list of 1,697 concepts.

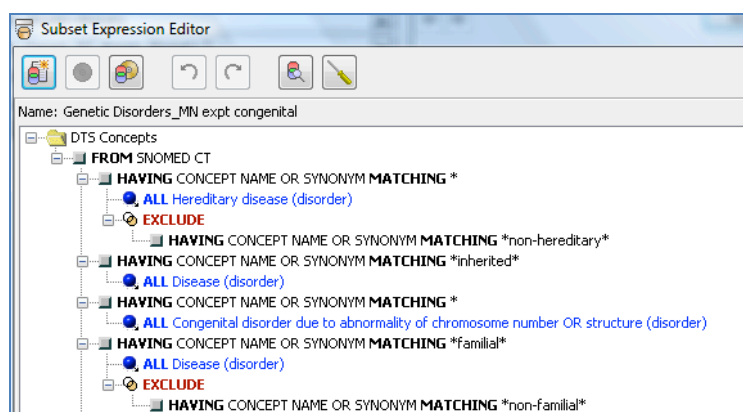


Figure 1. Partial representation of a lexical and semantic query of SNOMED CT to define a set of candidate Genetic Disorder concepts from SNOMED CT

To cross-check our results, we compiled a list of disorders from the Genetics Home Reference website hosted by the US National Library of Medicine (GHR, <http://ghr.nlm.nih.gov/>). We identified 604 distinct disorders in GHR. We successfully mapped 487 of these disorders back to SNOMED CT and expanded the result to include all their hierarchical descendants, resulting in a total of 1366 concepts.

The correspondence between the two lists was surprisingly low, with only 756 common concepts. There were 610 concepts found by mapping from GHR that were not found by our Boolean query, and 941 concepts in the Boolean query result that were not mapped from a GHR disorder.

3. Conclusion

Complex terminologies such as SNOMED CT must be used carefully to ensure that the correct terms are available to users. This exercise has shown that creating an appropriate subset is more complex than it might appear at first, and that even the largest terminologies have coverage gaps.

References

- [1] http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/Press_Releases/20112701-IHTSDO-NLM-press.pdf