Identifying Missing Finding Site Relations in SNOMED CT

Jonathan P. Bona, PhD[1], Werner Ceusters, MD[1],
[1]Department of Biomedical Informatics, University at Buffalo, Buffalo, NY

Introduction

Every concept in SNOMED CT comes with descriptions, including a *Fully Specified Name (FSN)*. FSNs typically end with a *semantic tag* in parentheses to disambiguate from other concepts with similar descriptions. However, some things that are not semantic tags appear in parentheses as parts of FSNs, and some FSNs have more than one tag, so it's sometimes hard to know by looking at an FSN exactly what its tags are. We will call words within parentheses in FSNs *apparent tags* if it's not obvious that they are *not* tags. "*(bullous edema is not sufficient evidence to classify a tumor as T4),*" appears in an FSN but is clearly not a semantic tag. Neither is *"(s)."*

SNOMED CT's concept model includes associative relations between concepts used to formally specify their meanings. For example, **10000006 | Radiating chest pain (finding)** is associated via **363698007 | Finding site (attribute)** with **51185008 | Thoracic structure (body structure)**. Formally specified and query-able relations between concepts enhance the completeness and usefulness of an ontology, and increase its utility for organizing and retrieving data, e.g., as part of structured EHRs, especially when combined with the Is-a hierarchy and subjected to logical inference. SNOMED CT's main finding concept **404684003 | Clinical finding (finding)** has many thousands of sub-concepts. Here we examine the use and disuse of finding sites in *finding* concepts related to cancer staging (specifically, some of the 1038 concepts under **385356007 | Tumor stage finding (finding)**). For those lacking *finding sites*, we use apparent tags in the FSNs to find body structure concepts that should be associated with those findings. This approach is motivated by the observation that although there is some relationship between the use of a certain semantic tag and the hierarchy in which the concept appears, the exact relationship is nowhere specified.

Method

We first found all *Tumor stage finding* concepts that lack *finding site* relations to any *body structure*, and identified170 candidate semantic tags in their FSNs. These were manually filtered for likely body structure tags: apparent tags that name anatomical entities. We focused on the 47 simplest body structure tags and omitted tags with longer phrases (e.g. "liver, including intrahepatic bile ducts"), and 36 short "tags" like "pT1-pT4, pM1" that are codes for different components of tumor staging. Each of these 47 body structure tags was manually matched to the best fit body structure concept. Starting again from the full set of T*umor stage finding*s, we used the body structure tags and matching concepts to generate a list of suggested finding sites for any of the 1038 concepts that used those tags.

Results

The result is a list of suggested candidate finding sites for 369 concepts (out of 1038) that do not have associated finding sites. These were manually verified as coherent and plausible finding sites for the relevant concept. Each concept was then marked with its nearest anatomically-specific TNM finding ancestor concept (e.g. **397092003 | Gallbladder TNM finding (finding)**) where applicable. Almost all had anatomically-specific TNM ancestors. Of the 17 concepts without such ancestors, 6 have the apparent tag "appendix." Unlike other body structures (*Bone, Thyroid*, etc) there is no TNM finding concept for *Appendix*. 4 out of 17 are pN (regional lymph nodes) or pM (distant metastasis) breast cancer findings. Another was for a primary tumor that cannot be assessed. Since none of these is about things located in the breast, they might need no finding site relation with Breast (body structure). The prevalence of anatomically-specific TNM finding concepts here suggests another technique to efficiently add finding sites for many of the tumor stage concepts: such concepts (**Ampulla of Vater TNM Finding**, etc) should be directly associated with the relevant body structures (**Structure of ampulla of**) -- relations that will apply also to their descendants (**pT1: Tumor limited to ampulla of Vater …**) after classification.

Conclusion

Our preliminary investigation of the use of *finding site* associations in a small section of the SNOMED hierarchy reveals an important gap that limits the usefulness of the relevant concepts for representing and reasoning about cancer staging and its relation to patients, parts of their bodies, their health, and so on. We have used the presence of apparent semantic tags that name body structures to align approximately one third of the target concepts with body structure concepts suitable to serve as their finding sites.