Title: Identifying Missing Hierarchical Relations in SNOMED CT from Logical Definitions Based on the Lexical Features of Concept Names

Presenter: Olivier Bodenreider, National Library of Medicine, USA

Audience
Developers and users of SNOMED CT interested quality assurance.

Objectives
1. To discuss quality assurance challenges in SNOMED CT;
2. To discuss how the lexical features of concept names reflect hierarchical relations;
3. To describe how description logics can be used to relate concept names.

Abstract

Background. Quality assurance of large biomedical terminologies remains an active area of research. For example, recent investigations of SNOMED CT have highlighted issues in its hierarchical structure and demonstrated their detrimental consequences (e.g., [1]).

Objectives. To identify missing hierarchical relations in SNOMED CT from logical definitions based on the lexical features of concept names, i.e., based on a bag-of-word approach to relating concept names.

Methods. We first create logical definitions from the lexical features of concept names, which we represent in OWL EL. We infer hierarchical (subClassOf) relations among these concepts using the ELK reasoner. Finally, we compare the hierarchy obtained from lexical features to the original SNOMED CT hierarchy. We review the differences manually for evaluation purposes.

Results. Applied to 15,833 disorder and procedure concepts from two sub-hierarchies, Disorder of head and Operative procedure on head, our approach identified 559 potentially missing hierarchical relations, of which 78% were deemed valid.

Conclusions. This work demonstrates that a bag-of-word approach to relating concept names can be easily implemented, leveraging description logic principles and tooling already used by SNOMED International. However, missing hierarchical relations are most often only the manifestation of incomplete or inaccurate logical definitions. While insufficient by itself, this simple, automated approach could guide the review of SNOMED CT content. More detail about this work are available in [2]. In the future, we will refine the approach by adding synonyms and lexico-syntactic analysis of the concept names, and apply it to the entirety of SNOMED CT.

References

1. Rector AL, Brandt S, Schneider T. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. J Am Med Inform Assoc. 2011 Jul-Aug;18(4):432-40.
2. Bodenreider O. Identifying missing hierarchical relations in SNOMED CT from logical definitions based on the lexical features of concept names. Proceedings of the 6th International Conference on Biomedical Ontology (ICBO 2016) 2016:(electronic proceedings: http://ceur-ws.org/Vol-1747/IT1601_ICBO2016.pdf).