Building an experimental German
user interface terminology linked to SNOMED CT

Presenter: Stefan Schulz, Medical University of Graz (Austria)

# Audience

1) SNOMED CT implementers in jurisdictions for which no official localized version exist.
2) Everybody interested in using SNOMED CT for NLP and text mining

# Objectives

To present an approach of large-scale, inexpensive and incremental creation of interface terms linked to SNOMED CT.

# Abstract

User interface terminologies (UITs) are collections of terms used in written and oral communications within a group of users. UITs depend on dialect, time, clinical specialty, and professional group. UITs acquire their semantic import by linkage to reference terminologies like SNOMED CT. Contrary to translations of reference terminologies, UITs have descriptive and not prescriptive character and therefore not part of a terminology standard. Both the English and Spanish release of SNOMED CT include aspects of UITs, as they provide synonyms besides FSNs and PTs. The EU project ASSESS CT (Assessing SNOMED CT for Large Scale eHealth Deployments) has recently recommended broad efforts to be invested into UITs [1].

One important use case for UITs is the provision of dictionary entries for natural language processing systems. Applying a semi-automated approach towards UIT building [2], the Medical University of Graz is addressing the needs for large-scale NLP-based annotations of clinical texts within CBmed – IICCAB [3]. Part of this is the creation of an experimental German user Interface terminology linked to the current international version for SNOMED CT. This UIT feeds the currently largest SNOMED CT use case in the German-speaking country.

The German UIT has currently 1.87 million entries. They are automatically created from a core vocabulary with 92,500 German short terms (including grammatical features) linked to 85,400 English short terms. For benchmarking the progress, we use random samples from a parallel corpus extracted from MEDLINE, constituted by publication titles in German and English. Currently the relative coverage by the German UIT amounts to 33.1% compared to 55.4% for the English descriptions in the international SNOMED CT release. The figures refer to strict token-level string match and need to be interpreted in the light of the morphologic richness of German medical language.

# References

1. Kalra, D., Schulz, S., Karlsson, D., et al. ASSESS CT Recommendations (2016).
<http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf>
2. Schulz, S. Using language technology for SNOMED CT localization" (44). SNOMED CT Expo 2015,
<https://confluence.ihtsdotools.org/pages/viewpage.action?pageId=12780196>
3. Schulz S. Innovative Use of Information for Clinical Care and Biomarker Research. <http://goo.gl/wHMedz>