

Extracting and representing clinical knowledge using SNOMED CT

Automatic indexing of clinical concepts in free text patient records using a standard medical terminology will enhance semantic retrieval, which can then be used for important applications such as decision support and disease outbreak detection. SNOMED CT is a rich terminology that provides standardisation of knowledge and language in the clinical domain. Two important challenges are identification of the concepts in clinical reports and then using the identified concepts to construct an integrated representation of the patient case. Although most clinical words found in the patient notes are present in the terminology, the rich set of relationships between the words and concepts cannot be fully represented. Lexical and concept verification is error prone due to the variance of the clinical language used in different departments in hospitals and the ungrammatical nature of the narrative reports. To integrate the recognised concepts extensions need to be made to the ontology or it needs to be placed in a wider ontological model to fully represent all matters relevant to the patient case.

This research aims to address the concept extraction and concept representation issues, by classifying medical concepts into the 17 SNOMED CT semantic categories, and representing their relationships using SNOMED CT 60+ relationship categories. The experiments will be conducted on a subset of a 44 million token Intensive Care corpus from the Royal Prince Alfred Hospital, Sydney. Through the classification, an extended ontology will be built to represent the ICU terms for use in data retrieval activities.