

SNOMED Clinical Terminology Mapping from Free-text Histopathology and Radiology Reports

Anthony Nguyen¹, Michael Lawley¹, David Hansen¹ and Shoni Colquist²

¹The Australian e-Health Research Centre, CSIRO ICT Centre, Brisbane, Australia

²Queensland Cancer Control Analysis Team, Queensland Health, Brisbane, Australia

Abstract

Objective: A system to automatically map free-text medical reports into the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) ontology is proposed. The mapping provides a standard reference terminology and a means to aggregate clinical information for retrieval and analysis of healthcare systems and processes.

Background: Free-text medical reports are non-standardised and make it difficult to aggregate clinical information for retrieval or analysis. Standardising the text into a standard reference terminology can overcome this problem. SNOMED-CT [1] is a large formal ontology of clinical terms that has been indentified by the Australian National e-Health Transition Authority (NeHTA) as the standard set of clinical terms to be used in systems within Australian healthcare [2]. The comprehensiveness and inherent structure of SNOMED-CT makes it suitable as a standard reference terminology.

Methods: The SNOMED-CT mapping tool was built to standardise free-text medical reports. The mapping was generated through a process of detection of negation phrases, extraction of Unified Medical Language System (UMLS) concepts (those restricted to the SNOMED-CT source) using the MetaMap Transfer (MMTx) application [3] to match phrases within sentences to the closest UMLS concepts, and then mapping the UMLS concepts to SNOMED-CT concepts using a look-up table. Post-processing of the SNOMED-CT concepts were performed to replace non-current concepts with their current forms and transform the resulting concepts to their long normal forms so that each concept is represented by its complete definition.

Results: The free-text to SNOMED-CT mapping system generates a standardised text of SNOMED-CT concepts which can be used for retrieval or analysis. Negation phrases detected can be used to negate relevant SNOMED-CT concepts, and clinical information can be retrieved using relevant queries. The SNOMED-CT mapping system has been applied to histopathology and radiology reports and is conjectured to provide increased knowledge to complement the cancer staging system developed in [4].

[1] College of American Pathologist, "SNOMED CT Clinical Terms Users Guide," January 2007.

[2] National E-Health Transition Authority, "Fact sheet: A national approach to clinical terminologies", August 2006, Available from <http://www.nehta.gov.au/>.

[3] U.S. National Library of Medicine, "MetaMap Transfer (MMTx)," 2008. Available from <http://mmtx.nlm.nih.gov/>, 2008

[4] I. McCowan, D. Moore, A. Nguyen, R. Bowman, B. Clarke, E. Duhig and M, Fry, "Collection of Cancer Stage Data by Classifying Free-text Medical Reports," *Journal of the American Medical Informatics Association* 14(6), 736-745, Nov/Dec, 2007.