Chapter 13
Automated Generation of SNOMED CT Subsets from Clinical Guidelines

Carlos Rodríguez-Solano
University of Alcalá, Spain

Leonardo Lezcano
University of Alcalá, Spain

Miguel-Ángel Sicilia
University of Alcalá, Spain

ABSTRACT

Recently, there has been a growing body of literature on how the large SNOMED CT (SCT) terminology could be implemented and used in different clinical settings. Its complexity and size is a major impediment for coding clinical information in practical applications. Therefore, it is sometimes necessary to define subsets for various use cases and specific audiences. Subsets are clusters of SNOMED CT terms that share a specified common characteristic. The automated generation of subsets from clinical document corpora have been proposed elsewhere, but they still require a collection of documents that is representative for the targeted domain. In this chapter the authors extend the research described in Rodríguez-Solano, Cáceres, and Sicilia (2011), where clinical guidelines’ glossaries are used as seed terminologies to automatically generate subsets by traversing SNOMED relationships. In the current research, further results have been obtained considering additional clinical guidelines; the application of quantitative analysis to the generated Snomed CT subsets, derived as result of implementing the proposed techniques, has allowed the evaluation of them.

DOI: 10.4018/978-1-4666-3667-5.ch013
INTRODUCTION

Semantic Interoperability between heterogeneous healthcare systems requires improvements in the precision of meaning and understandability during the exchange of information; in particular, for Electronic Health Record (EHR) systems, it is essential to enable the consistent use of clinical terminologies.

SNOMED CT (Systematized Nomenclature of MedicineClinical Terms) (http://www.ihtsdo.org/snomed-ct/) is one of such terminologies, whose adoption is fostered by the IHTSDO (International Health Terminology Standards Development Organisation) with a worldwide scope. In the last few years, has been promoted as an standard reference terminology aimed at achieving a means of standardised recording of medical information, and the interoperability between clinical systems.

OpenGALEN (http://www.opengalen.org) is a not-for-profit organization that provides another medical terminology. The GALEN Common Reference Model, is the model of medical concepts (or clinical terminology) being built in a formal language called GRAIL (GALEN Concept Representation Language). Currently available open source resources include a sophisticated ontology development environment for the medical domain.

The International Statistical Classification of Diseases and Related Health Problems (known as ICD) is a medical classification list for the coding of diseases, signs, and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases, as maintained by WHO (World Health Organization) (http://www.who.int/classifications/icd).

One of the major advantages of the recent SNOMED releases is that the structured nature of the vocabulary allows an ability to map from SNOMED to ICD 9 and, for this purpose, files are provided. Generally the co-occurrence of a number of SNOMED concepts can imply an ICD diagnosis.

Another extensive terminology is the Logical Observation Identifier Names and Codes database (LOINC) (http://loinc.org/), which is a database and universal standard for identifying medical laboratory observations. It was developed in 1994 and is maintained by the Regenstrief Institute, a US non-profit medical research organization. LOINC was created in response to the demand for an electronic database for clinical care and management.

The Unified Medical Language System (UMLS) (http://www.nlm.nih.gov/research/umls/) is a compendium of three knowledge sources in the biomedical sciences; it may also be viewed as a comprehensive thesaurus and ontology of biomedical concepts One of them is the UMLS Metathesaurus, a large, multi-purpose and multilingual vocabulary database that currently comprises more than 1.5 millions biomedical terms from over 100 sources.

Although SNOMED CT (SCT) has gained adoption in the last years, as previously mentioned, its practical application for coding clinical information is hampered by its complexity and size. It contains over 300,000 concepts in the latest release, and over 1 million relationships between these concepts; therefore, the large scale of SNOMED CT is a major barrier to its progress while there is evidence that only a small fraction of its content is being used (Stroetmann et al., 2009).

Because of the very wide range of concepts, and the rate of change of those concepts (Spackman K., 2005), dealing with the entire set of terms contained in SNOMED CT is difficult. In particular, when clinicians use SNOMED browsers to code clinical documents, experiments have shown their performance to be disappointing (Chiang, et al., 2006). Moreover, these ontologies that sometimes contain more than a hundred thousand concepts (terms) are hard to maintain as changes can affect large parts of the model.

Therefore, extracting meaningful fragments from a large terminology is a key issue for using
Related Content

Online Communication and Healthcare: The Diffusion of Health-Related Virtual Communities
www.igi-global.com/chapter/online-communication-healthcare/40655?camid=4v1a

eICU STUDY: A Proof of Concept
www.igi-global.com/article/eicu-study-proof-concept/115529?camid=4v1a

Confirmatory Factor Analysis to Establish Determinants of Wireless Technology in the Indian Healthcare
www.igi-global.com/chapter/confirmatory-factor-analysis-establish-determinants/50160?camid=4v1a

An Approach to Participative Personal Health Record System Development
www.igi-global.com/article/approach-participative-personal-health-record/53870?camid=4v1a