2 Introduction

Background

SNOMED CT is a clinically validated, semantically rich, controlled terminology. SNOMED CT is comprised of meaning-based concepts, human-readable descriptions and machine-readable definitions. SNOMED CT is used within electronic health records to support data capture, retrieval, and subsequent reuse for a wide range of purposes. SNOMED CT is also used to enable or enhance analysis of patient records and other clinical documents containing no original SNOMED CT content.

SNOMED CT hierarchies and formal concept definitions allow selective information retrieval to support analysis — from patient-based queries to operational reporting, public health reporting, strategic planning, predictive medicine and clinical research. As the SNOMED CT encoding of healthcare data increases, so too have the benefits being realized from analytics processes performed over this data.

Purpose

The purpose of this document is to review current approaches, tools and techniques for performing data analytics using SNOMED CT and to share developing practice in this area. It is anticipated that this report will benefit members, vendors and users of SNOMED CT by promoting a greater awareness of both what has been achieved, and what can be achieved by using SNOMED CT to enhance analytics services.

Scope

This document presents different data approaches, tools, terminology techniques, query languages, data architectures and user interfaces that may be used in performing analytics using SNOMED CT. Analytics services considered include patient-based queries, operational reporting, the application and audit of evidence-based medical practice, strategic planning, predictive medicine, public health reporting and clinical research. The benefits and challenges of these approaches are also presented. The case study summaries describe a selection of SNOMED CT analytics projects and tools.

This document does not provide an exhaustive list of analytics projects and tools, and does not mandate a specific approach. The development of clinical case definitions is also outside of the scope of this document.

Audience

The target audience of this document includes:

- Members who wish to learn about current analytics activities in other jurisdictions and inform future directions;
- Clinicians, informatics specialists and technical staff involved in the planning, management, design or implementation of clinical record applications or healthcare analytics tools;
- Software vendors, data analysts, epidemiologists and others designing SNOMED CT based solutions.

This document assumes a basic level of understanding of SNOMED CT. For background information it is recommended that the reader refers to the SNOMED CT Starter Guide.

Document Overview

This document presents an introduction to analytics over data with SNOMED CT content.

Section 1 (Executive Summary) provides a concise summary of the document.

Section 3 (Introduction) introduces the document by explaining the background, purpose, scope, audience and overview of the document.

Section 4 (SNOMED CT Overview) introduces the topic by presenting a definition of analytics and describing the scope, purpose and substrates of analytics services which use SNOMED CT.

Section 5 (Preparing Data for Analytics) describes some approaches used to prepare clinical data for analytics using SNOMED CT, including mapping and natural language processing.

Section 6 (SNOMED CT Analytics Techniques) presents a range of techniques for using SNOMED CT to perform data analytics, including using value sets, subsumption, defining relationships and description logic.

Section 7 (Task-Oriented Analytics) looks at how these SNOMED CT based techniques can be used to assist with specific analytics tasks for point of care analytics, population health monitoring and reporting, and clinical research.

Section 8 (Data Architectures) presents a number of approaches for architecting analytics services, including querying directly over patient data, using a data warehouse, querying a virtual medical record and using distributed storage and processes.

Section 9 (Database queries) considers the query languages that are needed to perform analytics over the combination of the patient record and terminology content.

Section 10 (User Interface Design) presents a selection of user interface styles that may be used with SNOMED CT to support querying and results visualization.
Section 11 (Challenges) discusses some of the challenges which are faced when performing analytics over SNOMED CT enabled data, including the reliability of patient data, information model/terminology boundary issues, concept definition issues, versioning and inactive content.

Two appendices to this report present a variety of project case studies and vendor tooling case studies respectively. These appendices, which are referenced extensively throughout this document, can be found at [http://snomed.org/analyticscasestudies](http://snomed.org/analyticscasestudies).