

6 SNOMED CT Analytic Techniques

SNOMED CT offers a number of analytics techniques, which are not possible using other coding systems. SNOMED CT's hierarchical design improves upon the purely lexical query capabilities of free text lists or 'flat' controlled vocabularies. For example, a purely text based query for 'kidney disease' will not return the kidney disease 'glomerulonephritis'. Purely mono-hierarchies, however, limit querying to a single grouping of each code. For example, using a mono-hierarchy 'tuberculosis of the lung' must be assigned a code which makes it *either* a kind of 'lung disease' *or* a kind of 'tuberculosis' – however it cannot be both. Using SNOMED CT's polyhierarchy 'tuberculosis of the lung' can be represented as both a kind of 'lung disease' *and* a kind of 'tuberculosis'. The inclusion of other attribute-based defining relationships and the ability to represent SNOMED CT using OWL 2 EL enables additional Description Logic techniques for classifying and querying SNOMED CT. Extending these capabilities even further, it is possible to use Description Logic techniques across both the terminology and the structure of the patient records in which the codes are stored. Finally, in some specific use cases such as billing, reimbursement and statistics where double counting must be avoided, clinically recorded SNOMED CT codes can be used to map into more general statistical classifications, such as ICD (International Classification of Diseases).

In this section, we describe how the following analytics techniques can be used to support analytics over SNOMED CT enabled data. The techniques described include:

- Subsets – for example, find the patients with a diagnosis in the set of 'kidney disease codes'
- Subsumption – for example, find the patients with a diagnosis that is a subtype (or self) of 'kidney disease'
- Using defining relationships – for example, find the patients whose diagnosis has a finding site of 'kidney structure' (or a subtype of 'kidney structure')
- Description logic over terminology – for example, find the patients whose diagnosis is associated (directly or indirectly) with the 'Streptococcus pyogenes organism'
- Description logic over terminology and structure – for example, find the patients with a family history of heart disease (where this may either be recorded as 275120007 |family history: cardiac disorder| or recorded in a 'Family History' section on a form as 56265001 |heart disease|)
- Using statistical classifications – for example, to meet national reporting guidelines using ICD (International Classification of Diseases)

In practice, a query language may combine a number of these techniques in the same syntax. With the possible exception of the last two approaches, these SNOMED CT query techniques should then be embedded within an EHR query to ensure that the semantic context of the surrounding patient record is taken into account.
