

6.1 Subsets

One approach to analytics using SNOMED CT is to construct subsets of SNOMED CT identifiers, which are applicable to a specific clinical purpose, and to test the codes recorded in patient records to check for membership in the appropriate subset. Subsets of SNOMED CT identifiers may either be defined extensionally or intensionally.

Extensionally defined subsets are those in which each concept is individually enumerated. They are usually manually constructed and maintained, and can therefore be labor intensive and error prone. For example, one might construct a subset of kidney disease codes including 36171008 |glomerulonephritis|, 71110009 |hydrocalycosis| and 42399005 |renal failure|.

Intensionally defined subsets are those which are automatically populated (or expanded) based on a machine processable query. For example, one might construct a subset of kidney disease codes using the results of the query "<< 90708001 |kidney disease|" (i.e. descendants or self of 90708001 |kidney disease|). The query used to define an intensional subset may utilize SNOMED CT's hierarchical relationships, attribute values, descriptions, and membership in other intensionally or extensionally defined subsets. For more information about SNOMED CT query languages, which may be used to define intensional subsets, please refer to section 9 Database Queries.

Example

A subset containing types of 58437007 |tuberculosis of meninges| may be defined extensionally as follows:

Concept ID	Description
58437007	tuberculosis of meninges (disorder)
90302003	tuberculosis of cerebral meninges (disorder)
38115001	tuberculosis of spinal meninges (disorder)
11676005	tuberculous leptomeningitis (disorder)

With the help of the SNOMED CT hierarchy (as shown in Figure 6.1-1), this same subset can be defined intensionally as:

```
<< 58437007 |tuberculosis of meninges|
```

The expansion of an intensional subset defined using this query is the same as the extensionally defined subset shown above.

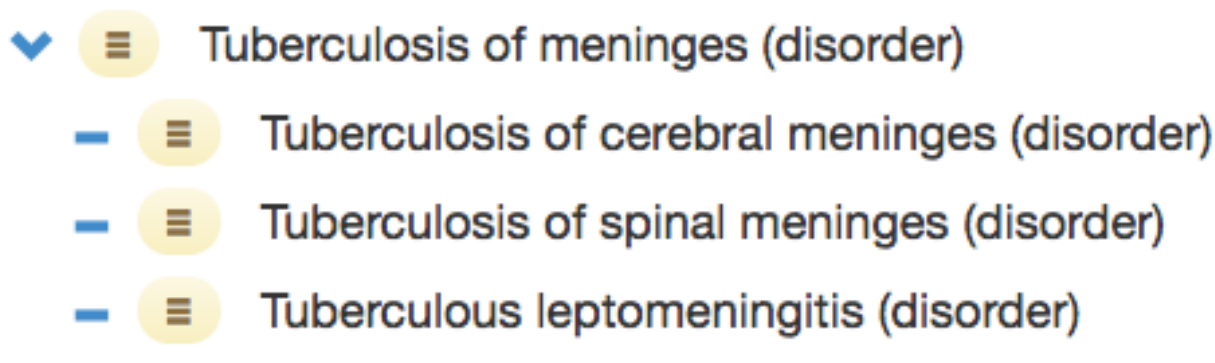


Figure 6.1-1: Tuberculosis of meninges concept sub-hierarchy

Using a lexical query, it is also possible to intensionally define a subset of 'tuberculosis of meninges' findings. However, the results of purely lexical queries are not as reliable. For example, using the query:

```
<< 404684003 |clinical finding| {{ term = ".tuberculosis.*meninges." }}
```

the following expansion can be calculated:

Concept ID	Description
58437007	tuberculosis of meninges (disorder)
90302003	tuberculosis of cerebral meninges (disorder)
38115001	tuberculosis of spinal meninges (disorder)

As can be seen, the results of this lexical query only includes 3 of the possible 4 values from the previous subset. In other cases, lexical queries may incorrectly find concepts which are not appropriate to the subset. It is therefore recommended that lexical queries are avoided in the definition of intentional subsets. However, they do serve a useful purpose in identifying candidates for an otherwise manually crafted subset.

Implementation

Defining Subsets in SNOMED CT

Subsets of SNOMED CT may be defined locally as a flat list of concept identifiers, or as an independent query specification. However, where wider distribution and/or version control is required over these subsets, SNOMED CT reference sets offer the ideal solution.

Extensional subsets are commonly defined in SNOMED CT using a Simple reference set - however an Ordered reference set or Annotation reference set can be used if additional information needs to be recorded for each member of the subset. Intensional subsets are defined in SNOMED CT using a Query specification reference set. A Query specification reference set allows a serialized query to define the membership of a subset of SNOMED CT components. It also specifies the extensional reference set into which the results of executing the query are generated. Intensional reference sets are preferred in many circumstances as they enable their membership to be automatically recomputed over new versions of SNOMED CT. Version management of subsets is discussed further in section [11.4 Versioning](#).

Subsets can be created using the following methods, either alone or in any combination:

- Manual inclusion, using search and browse methods
- Existing subset, used as a starting point for further manual inclusion and update
- Lexical queries, to identify candidate members, followed by manual verification and update
- Hierarchical queries, to identify descendants of a given concept (e.g. descendants of <73211009 |diabetes mellitus|)
- Attribute queries, to identify concepts with a specific attribute value (e.g. disorders with a finding site of 80891009 |heart structure|)
- SNOMED CT queries, using the SNOMED CT Expression Constraint or Query languages, which offer additional query functionality. Please refer to section [9 Database Queries](#) for more details.

Case Studies

A number of vendor products, such as [Apelon](#) and [B2i Healthcare](#) allow users to create customized extensional and intensional subsets of SNOMED CT. Other vendor products, such as the [Cambio COSMIC®](#) Electronic Patient Record system, [Caradigm's](#) population health solutions, [Cerner's](#) data warehousing solution and [Epic's](#) decision support and reporting tools use subsets of SNOMED CT to support their analytics services.

¹ Expansion derived from SNOMED CT International Edition, dated 20170131.