

6.4 Description Logic Over Terminology

SNOMED CT's semantics are based on Description Logic (DL). This enables the automation of reasoning across SNOMED CT, and subsequently the implementation of more powerful analytics operations than is possible using most other approaches. In addition to the subsumption and defining relationship testing described in the previous approaches, DL reasoners and query engines are able to utilize a number of additional logic-based techniques including:

- Property chaining

A property chain is a rule that allows you to infer the existence of a property from a chain of properties. For example, "x has parent y" and "y has parent z" implies "x has grandparent z" (which may be written as "|has parent||has parent||has grandparent|"). The current release of SNOMED CT includes the property chain:

```
363701004 |direct substance| 127489000 |has active ingredient|  
363701004 |direct substance|
```

However, more property chains may be added in local implementations if required.

- Reasoning over concrete values

Some concepts in SNOMED CT (e.g. 374646004 |amoxicillin 500mg tablet|) require numbers or strings to fully define their meaning. By generating an OWL 2 representation of these concept definitions, Description Logic can be used to reason over their complete definition (including the concrete values)

- Testing equivalence and subsumption of postcoordinated expressions (without calculating normal forms)

Description Logic enables equivalence and subsumption testing to be performed efficiently, without the need to manually calculate the normal form of each expression.

- Reasoning over minimum sufficient sets

SNOMED CT definitions include the set of necessary and sufficient conditions that define the given concept. However, SNOMED CT does not currently distinguish the minimum sets which are sufficient to define these concepts. For example, the defining relationships of 154283005 |pulmonary tuberculosis| are:

```
116680003 |is a| = 64572001 |disease|  
246075003 |causative agent| = 113858008 |mycobacterium tuberculosis complex|  
116676008 |associated morphology| = 6266001 |granulomatous inflammation|  
363698007 |finding site| = 39607008 |lung structure|
```

However, while the associated morphology of 'granulomatous inflammation' is necessarily present, the following set of defining relationships are sufficient to infer 154283005 |pulmonary tuberculosis|:

```
116680003 |is a| = 64572001 |disease|  
246075003 |causative agent| = 113858008 |mycobacterium tuberculosis complex|  
363698007 |finding site| = 39607008 |lung structure|
```

Using Description Logic, it is possible to reason using multiple minimum sufficient sets for each concept.

Example

For example, if we want to find all disorders that are associated with the organism 80166006 |streptococcus pyogenes|, we may discover (using the SNOMED CT Relationships file) that there is a direct 'causative agent' relationship from 302809008 |streptococcus pyogenes infection| to 80166006 |streptococcus pyogenes|. However, by introducing the following property chain rule:

```
47429007 |associated with| 47429007 |associated with| 47429007 |associated with|
```

and noting that 47429007 |associated with| has three subtypes:

```
255234002 |after|
```

```
42752001 |due to|
```

```
246075003 |causative agent|
```

it is possible to discover, using Description Logic, that 81077008 |acute rheumatic arthritis| and 58718002 |rheumatic fever| are also 'associated with' the concept 30209008 |streptococcus pyogenes infection|. Figure 6.4-1 illustrates these relationships that can be discovered using property chaining.

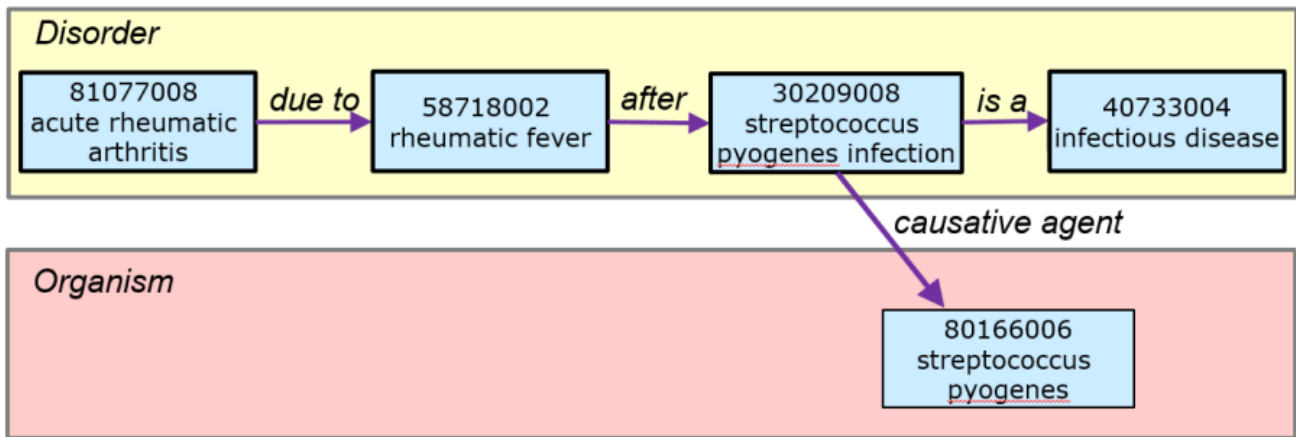


Figure 6.4-1: Property chaining

Implementation

OWL 2

Using Description Logic techniques to perform analytics over SNOMED CT involves first translating SNOMED CT into OWL 2 (Web Ontology Language). OWL 2 is an ontology language for the Semantic Web with formally defined meaning. The SNOMED CT international release comes with a Perl transform script that converts the RF2 files into OWL XML/RDF, Functional Syntax or KRSS files.

Once generated, the OWL files can then be loaded into a Description Logic Editor (such as Protégé) or used directly by a terminology service which offers description logic capabilities. The Description Logic Editor or terminology service then uses DL reasoners (also known as 'classifiers'), such as Snorocket, ELK and FACT++, to perform consistency checking and subsumption testing (also known as 'classification') over SNOMED CT. Subsumption testing can also be performed between two expressions. Semantic query languages, such as SPARQL, can be used to query over RDF representations of SNOMED CT.

Case Studies

Some commercial terminology servers, such as [B2i Healthcare's](#) Snow Owl terminology server, use Description Logic based techniques to support both classification and querying over SNOMED CT. [Kaiser Permanente](#) is collaborating with Oxford University to investigate ways of performing complex queries efficiently across extremely large numbers of patient records using scalable parallel processing and description logic reasoners.