6.5 Description Logic Over Terminology and Structure

When performing analytics over patient data, an appreciation for the semantics represented in both the terminology and the information model is required. Different information models can use different amounts of precoordination in the terminology, and the same semantics can be represented using different information structures. By using description logic over both the terminology and the information structures, a consistent representation of the meaning of data can be achieved, irrespective of whether this meaning is captured in the data values or in the model itself.

Example

Consider for example the two alternative ways of recording family history, as shown in Figure 6.5-1. The green rectangles represent the logical structure of the information model and the blue rectangles represent the concept identifiers that are used to populate this information model in the patient record.

The information model on the left uses a heading of 'Family history' to indicate that the named problem refers to a family history of that problem. The information model on the right uses the terminology value to indicate that the problem refers to a family history instance.
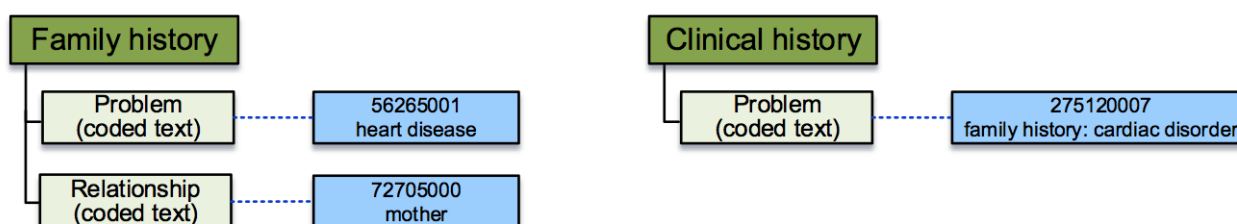

Figure 6.5-1: Two ways of recording family history

When querying over data, which may be collected in either format, both the semantics of the information model and the semantics of the data instances must be considered. One way of achieving this is to use an 'expression template' to convert all data instances into a Description Logic representation, and use this to reason over the data. Figure 6.5-2 shows an example of an expression template that could be used to create a SNOMED CT expression for each of the data instances shown in Figure 6.5-1. Please note that the orange parallelograms represent 'slots' which are subsequently populated with the value of the named data element (e.g. '$Problem').
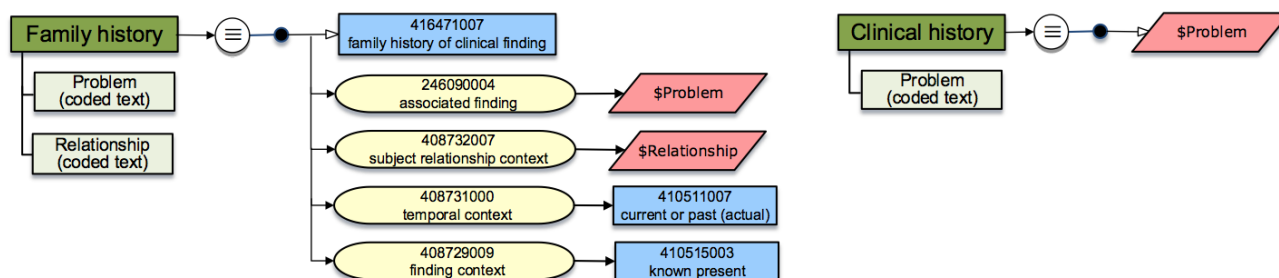

Figure 6.5-2: SNOMED CT expression representation of family history data

When the data instances from Figure 6.5-1 are used to populate the templates from Figure 6.5-2, the following two expressions are created:

416471007|family history of clinical finding|:

246090004 |associated finding| = 56265001 |heart disease|,

408732007 |subject relationship context| = 72705000 |mother|,

408731000 |temporal context| = 410511007 |current or past (actual)|,

408729009 |finding context| = 410515003 |known present|

275120007 |family history: cardiac disorder|

These expressions may then be compared using a DL reasoner to discover that the first expression is subsumed by the second, or queried using a semantic query language to allow the two data representations to be analyzed in a consistent way.

Implementation

OWL 2

Description Logic techniques, such as those described in section 6.4 Description Logic Over Terminology, can be used to reason over both the terminology and the information model. In addition to translating SNOMED CT to OWL 2, OWL 2 representations of the information model are also created using 'templates' that include 'slots' which are then filled with the patient record instance values. DL reasoners, such as Snorocket, ELK and FACT++, and semantic query languages, such as SPARQL, can then be used over both the terminology and the information model in a consistent way.

Case Studies

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. In this project, the analysis is being performed over an OWL-RL representation of the patient data, which incorporates both the terminology and the structure of the information.