SNOMED CT Concept Model
Yongsheng Gao, Romin Khazai

IHTSDO – International Health Terminology Standards Development Organisation

Abstract
Concept model specifies the definition of SNOMED CT concepts based on formal Description Logic. This paper provides an overview of the key aspects of the concept model, including visual diagrams and examples of language syntaxes.

Keywords:
SNOMED CT; Terminology; Logic

Introduction
SNOMED CT is one of the most comprehensive clinical terminologies developed and maintained by IHTSDO. The international content of terminology is released twice a year in January and July with new additions and updates. The latest January 2015 release includes a total of 417,209 concepts, 1,243,071 descriptions, 730,242 stated relationships and 2,455,945 inferred relationships.

A concept represents a clinical idea, action, or entity in SNOMED CT. There are three types of descriptions. The fully specified name is an unique and unambiguous term with a semantic tag such as (disorder) for a concept. A concept has a preferred term for a given language and different synonyms.

Given there are a large number of relationships between concepts, it is challenging to maintain them without a concept model, in particular for implicit relationships that can be inferred. The concept model specifies the logical definition of concepts based on the formal DL (description logic). The concept model can be seen in any SNOMED CT browsers, and the IHTSDO browser can be accessed at http://browser.ihtsdotools.org.

Key aspects of the concept model
Hierarchies of SNOMED CT
The terminology content is arranged in 19 hierarchical structures that covers a wide range of terms for clinical practices. The top-level concepts of these hierarchies are direct subconcepts of the root concept ‘SNOMED CT Concept’. The top-level concepts are disjoint, which means there is no shared concept between any of these hierarchies. For example, a concept can be either a procedure or finding, but not both. However, SNOMED CT is poly-hierarchical within a hierarchy, which is different to a classification system such as ICD (International Classification of Diseases). For example, within the procedure hierarchy, the concept ‘magnetic resonance imaging (MRI) guided biopsy’ can have two superconcepts of ‘biopsy’ and ‘MRI’. The following hierarchies are three examples that are commonly used:

Clinical finding/disorder hierarchy represents the result of a clinical observation, assessment or judgment of a patient. It includes both normal and abnormal findings, such as normal gait, bleeding from nose, asthma, and diabetes mellitus.

Procedure hierarchy represents a broad range of activities performed in the provision of health care, which includes diagnostic procedures (e.g. ultrasonography of abdomen), therapeutic procedures (e.g. kidney transplant), educational procedures, and administrative procedures (e.g. referral).

Situation hierarchy represents clinical findings, procedures or events in a specific context, possibility or occurrence of finding or procedure (e.g. suspected heart disease, absence of liver disease, planned kidney transplant procedure but not performed yet), temporal context (e.g. history of a disorder), and subject of record context that refers to a person other than the patient, e.g. family history of diabetes mellitus.

Attributes of SNOMED CT
SNOMED CT is a directed acyclic graph in mathematics and computer science. The attributes are directional and cyclic relationships are not allowed for the attribute ‘IS A’ that represents hierarchical relationship between two concepts. The ‘IS A’ relationship is a true subtype/supertype relationship and the overload of ‘member of’ and ‘part of’, which are not considered as true subtypes, are not allowed by following good practice for ontology construction.

Currently, there are 77 attributes for concept modeling. New attributes can be added and approved for improvement. The 7 history attributes, such as ‘SAME AS’, provide a link from inactive to active concepts for historical records.

Domain and range of attribute
The domain and range of an attribute specify constraints for a concept model. Domain is the scope of concepts that can have an attribute for concept modeling. Range specifies the values that an attribute can take. For example, attributes of ‘Finding site’ and ‘Associated morphology’ are allowed to be used for concept modeling for all clinical finding concepts and their ranges are morphology and body structures. Attributes like ‘Method’ and ‘Procedure site’ cannot be used for findings, but only procedures.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Attribute</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical finding</td>
<td>Associated morphology</td>
<td>&lt;&lt; Morphologically abnormal structure</td>
</tr>
<tr>
<td></td>
<td>Due to</td>
<td>&lt;= Clinical Finding</td>
</tr>
<tr>
<td></td>
<td>Finding site</td>
<td>&lt;= Anatomical or acquired body structure</td>
</tr>
<tr>
<td>Procedure</td>
<td>Direct morphology</td>
<td>&lt;&lt; Morphologically abnormal structure</td>
</tr>
<tr>
<td></td>
<td>Has intent</td>
<td>&lt;= Intents (nature of procedure values)</td>
</tr>
<tr>
<td></td>
<td>Method</td>
<td>&lt;&lt; Action</td>
</tr>
<tr>
<td></td>
<td>Procedure site</td>
<td>&lt;&lt; Anatomical or acquired body structure</td>
</tr>
<tr>
<td></td>
<td>Using substance</td>
<td>&lt;&lt; Substance</td>
</tr>
</tbody>
</table>
Role/attribute group and nesting

Attribute and role are used interchangeably in SNOMED CT. Role grouping has been introduced to achieve correct inferences for complex concepts that involve an identical attribute more than once in a concept model[2]. Two or more logically associated attribute/value pairs are grouped together. For example, ‘fracture dislocation of joint’ has two role groups: one represents the fracture of bone and the other is the dislocation of joint. They indicate which body site is associated with which morphological abnormality. Without the role grouping, an erroneous combination of dislocation of bone and fracture of joint may have occurred.

Role grouping is not an original construct from the DL and there are debates and confusions of its use and interpretation. It is effectively a nesting of expressions by an attribute named ‘Role group’. ‘Role group’ is represented by a numeric number of relationship groups in the current release format. This approach avoids nesting and recursive queries for relational databases. The disadvantage though is that nesting at more than one level cannot be properly represented, e.g. a role group cannot be within another role group.

The interpretation of ‘Role group’ can be ‘Has-part’ that takes condition or procedure expressed as an expression[3]. It could be more specific as ‘Has condition’ for clinical findings and ‘Has subprocedure’ for procedures. A Perl script is provided with the international release to convert SNOMED CT to OWL (Web Ontology Language)[4]. ‘Role group’ is represented as an attribute for nested expressions. The multiple levels of nesting will be allowed only if a specifically named attribute, such as ‘Has condition’, has been used.

Primitive vs. Defined

The ‘defining status’ indicates whether there is an equivalence between a concept model and its intended meaning presented by a fully specified name. A fully defined concept is expressed by necessary and sufficient characteristics. In contrast, a primitive concept is defined by necessary conditions only, for example, ‘Fracture of bone’ in the standard diagram representation[5]. The symbol ‘=’ in the circle divides the left and right sides of definition and the symbol indicates the equivalence. The black ‘dot’ represents conjunction of all statements at the right side. The left hand side ‘Fracture of bone’ is the concept that is fully defined by the right hand side ‘IS A’ relationship (open-headed arrow) to ‘Bone injury’ and a role group (empty circle) of two attribute/value pairs: attribute ‘Finding site’ and value ‘Bone structure’; attribute ‘Associated morphology’ and value ‘Fracture’.

It is important to recognize that the concept model definition is a logic one which must be true for all cases. The definition is not intended to represent clinical diagnostic criteria that can include symptoms, X-ray, or CT imaging which are not part of logical definition.

In figure 2, the symbol ‘⊂’ in the circle represents ‘Metabolic disease’ is subsumed by ‘Disease’ and both are primitive concepts.

Figure 2 – primitive concept model

The definition status has an impact to the DL classification results. The DL reasoner can detect equivalence and subsumption for fully defined concepts but not equivalence for primitive concepts.

Description Logic and reasoning services

SNOMED CT is based on the DL that is a family of knowledge representation formalism using attribute language to define meaning of terms. SNOMED CT conforms to the DL OWL 2 EL profile that is efficient for large bio-health ontologies[6]. The following are key axioms and constructs that have been implemented in SNOMED CT:

- Existential quantification, (SOME)
- Intersection of classes, (Conjunction, AND)
- Class inclusion (SubClassOf, IS A)
- Class equivalence (EquivalentClasses)
- Object property inclusion (sub-attribute)
- Domain restrictions
- Range restrictions
- Disjointness

Domain and range restrictions have been implemented technically in the IHTSDO terminology tooling to ensure correct concept modeling. The disjointness of the top level concepts are realized by editorial principles though this is not technically implemented. Some constructs are not implemented but might be in future SNOMED CT. For example, a new design of anatomy concept model has applied property chains, reflective and transitive object properties. The future pharmacy concept model can utilise the datatype property for concrete domains, e.g. numbers.

Constructs such as Universal quantification, Disjunction, Class negation, Inverse object properties are not part of the EL profile because of their complexity and high prerequisite on computation power and time. These constructs could be useful for SNOMED CT when the DL makes performance advancements in the future.

The DL reasoner known as Snorocket[7] has been used for SNOMED CT editing tool at IHTSDO and ELK[8] is used for the new anatomy design project because of their fast speed of classification for the OWL 2 EL profile.

The benefits for DL based design of concept modeling are that it allows formal logic based semantics for defining concepts. The auto-classification by the DL reasoner provides inference for detecting equivalence and subsumption relationships. This reasoning service can be demonstrated by stated and inferred views of a concept model. The stated view represents...
attributes and values of a concept definition stated by a modeler. The model is distributed in the ‘stated relationship table’ in the international release. The inferred view of concept definition are generated by the DL reasoner, which includes additional relationships and removes redundant relationships. The relationship table in the release is based on the inferred view. The different views are demonstrated by Kent Spackman’s diagrams with modifications.

![Figure 3 – Stated view example](image)

The fracture of femur is stated as direct subconcept of fracture of bone.

![Figure 4 – Inferred view example](image)

The inferred view is generated after the stated view has been classified by the DL reasoner. The differences between the two views can be easily noticed:

1. The subsumption between ‘femur fracture’ and ‘lower limb fracture’ is inferred by classifier because ‘femur’ is a subconcept of ‘bone structure of lower limb’.
2. Redundant relationship has been removed as a result of transitive reduction. ‘Fracture of femur’ is no longer presented as a direct subconcept of ‘fracture of bone’.

A transitive closure of ‘IS A’ relationship includes all relationships between fractures represented by solid and dotted arrows. The transitive closure is useful for intelligent queries, e.g. asking for all fractures of bone will return not only bone fractures, but also fractures of lower limb and femur fractures. IHTSDO has published a SQL script in technical implementation guide for generating a transitive closure table. It can also be generated by a programming language such as Perl, Java. A fast traversal of nodes via edges can be performed at runtime in a NoSQL graph database such as Neo4j, which can even eliminate the need for a static transitive closure table.

**Language syntaxes and Expressions**

The standard diagram is a user friendly graphic view of concept model. The communication and information exchange requires standard language syntax and expressions. IHTSDO has developed Compositional Grammar for representing concept model and expressions[9]. The OWL language is an international standard for representing ontologies in semantic network. SNOMED CT conforms to the OWL language standard.

A concept is represented by a single code, which is called pre-coordinated expression. For example:

73211009 [Diabetes mellitus], or concept identifier 73211009 on its own without description of ‘Diabetes mellitus’.

A concept is represented by a combination of codes according to the concept model, which is called post-coordinated expression.

The simplest form of post-coordinated expression is a combination of multiple focus concepts. For example, ‘Needle biopsy of kidney’ can be represented by an expression in compositional grammar:

7246002 [Kidney biopsy] + 129249002 [Needle biopsy]

The plus + represents conjunction of these two concepts and the new concept ‘Needle biopsy of kidney’ is a subconcept of them.

The most common usage of post-coordinated expression is to represent a new concept by refinement of an existing concept.

The post-coordinated expression can be represented by either Compositional Grammar or OWL. For example, radius fracture can be represented by an expression in compositional grammar:

125605004 [Fracture of bone]: 363698007 [Finding site] = 62413002 [Bone structure of radius]

The same concept can be represented in the Manchester OWL syntax[10].

‘Fracture of bone’ and ‘Finding site’ some ‘Radius’

The word and represents intersection of classes (conjunction) which is semicolon in compositional grammar. The word some represents existential quantification which is symbol ‘=’ in compositional grammar.

**Conclusion**

The concept model based on the formal Description Logic makes it possible to maintain such an enormous terminology with well organized hierarchical structures and logical definitions. The concept model conforms to the international standard of OWL 2 EL for maximum semantic interoperability, and also plays a critical role for machine-processing of meaning of concepts. The logical reasoning service provides great benefits for intelligent analysis of data based on conceptual meaning in addition to lexical description.

**References**


Ontology Language Primer (Second Edition), W3C Recommendation 11 December 2012.


[6] OWL 2 EL profile http://www.w3.org/TR/owl2-profiles/#OWL_2_EL


Address for correspondence
Yongsheng Gao, MB, PhD. Senior Terminologist, IHTSDO Gammeltorv 4, 1. 1457 Copenhagen, Denmark
yga@ihtsdo.org
tel: +44 (0)7508 509 219
http://www.ihtsdo.org