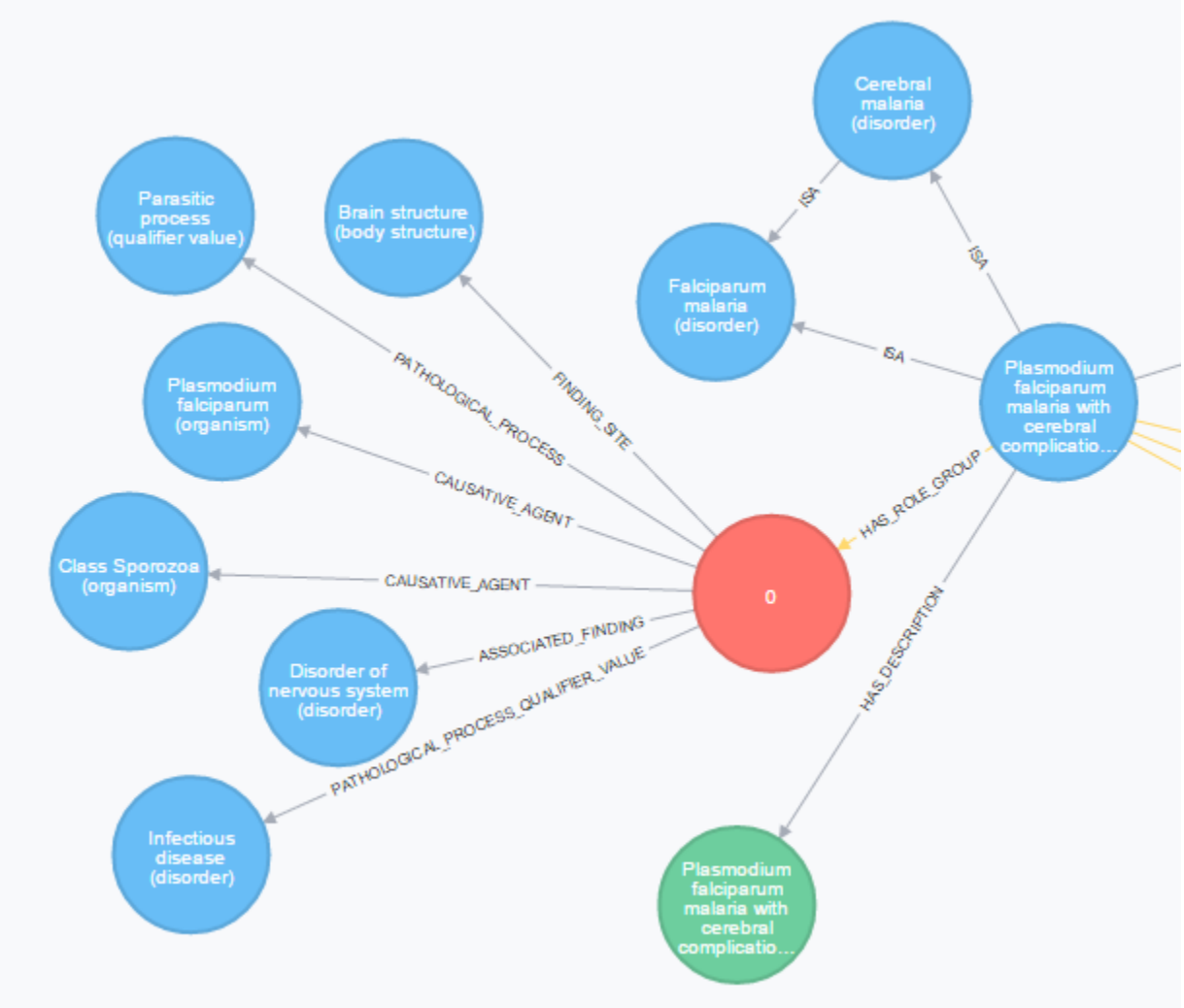
SNOMED CT Full View Graph Database Schema

Author: Jay Pedersen, University of Nebraska Medical Center, November 12, 2015

Document Version: 0.1
SNOMED_G version: 1.0



Introduction

This document describes the data structures stored within a NEO4J graph database to represent a full view release of SNOMED CT containing all of its history information including concepts, relationships, descriptions and role groups.

The goal of the database is to carry the entirety of the information in the Full View RF2 release files into a database, including all RF2-specified attributes for concepts, descriptions and relationships as well as the complete history of their changes. This design gives the ability to directly query based on the current values of any attribute, but also gives the ability to compute the value of any attribute at any time point in the SNOMED CT era.

Information Model Overview

Entities

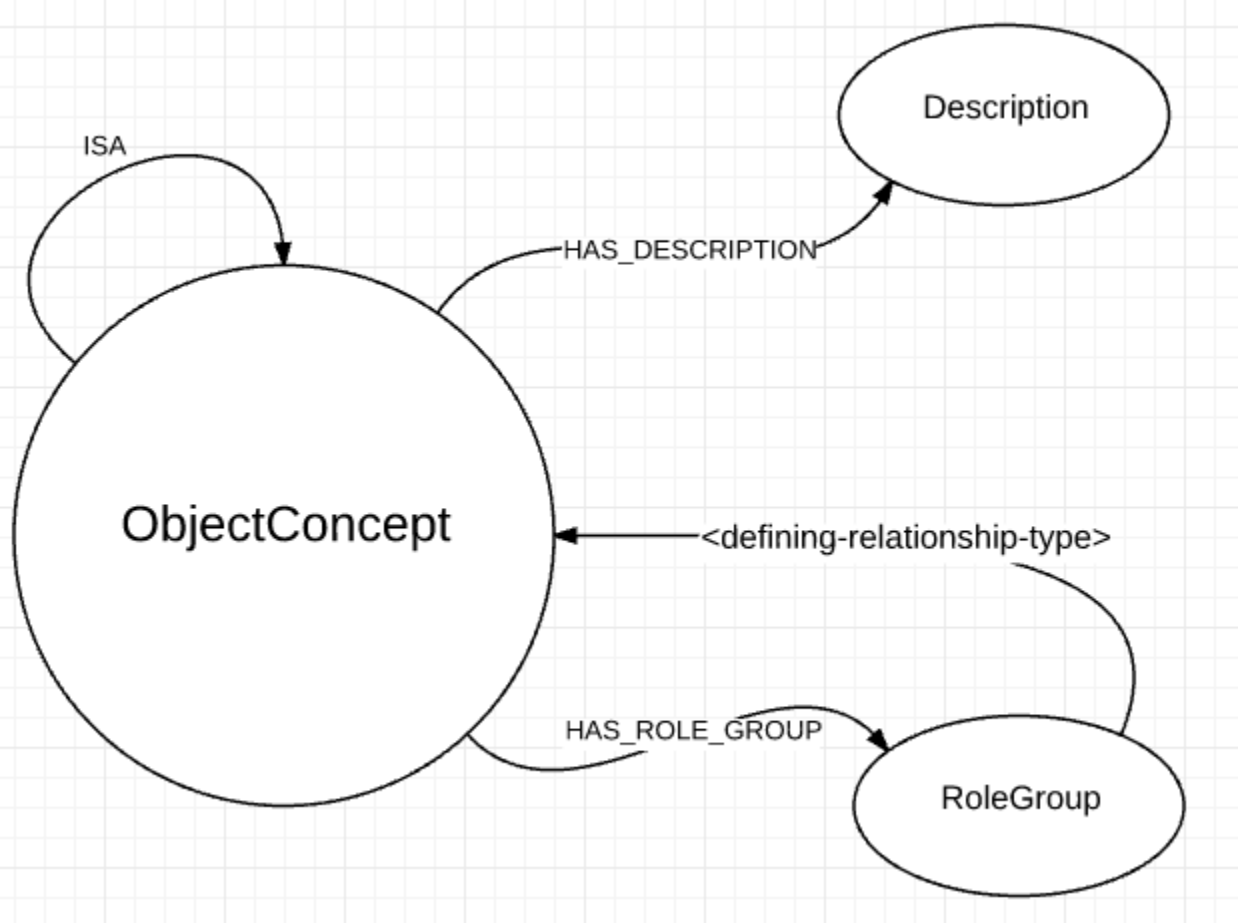
The following are the entities defined in the schema of a SNOMED_G graph database. A graph is composed of nodes and relationships (aka edges), and thus the entities are nodes and relationships between the nodes.

Entity Name	Type	Description
ObjectConcept	Node	A SNOMED CT concept with a unique SCTID identifier.
Description	Node	A textual description for a SNOMED CT concept, which may be its Fully Specified Name (FSN), its preferred term or a synonym.
RoleGroup	Node	A node used to hold definitional attributes for a SNOMED CT concept. There are one or more defining relationships (edges) associated with a RoleGroup; these yield a portion of the semantic definition of a concept. There can be multiple RoleGroup nodes associated with an ObjectConcept. A single RoleGroup is treated as an atomic semantic unit. The semantics of a SNOMED CT concept includes the union of semantics of all associated role groups, along with the semantics of concepts reachable by ISA relationships.
ISA	Edge	An ISA edge from ObjectConcept A to ObjectConcept B indicates that A is a subtype of B.
<defining-relationship>	Edge	These relationships are edges between RoleGroup nodes and ObjectConcept nodes. Each defines part of a role group. There are over 1,000 defining relationship types. For example, a concept for medical procedure will commonly have a role group which has a METHOD defining relationship. For example, the procedure “CT of petrous bones (procedure)” (SCTID 241522007) has a role group with a METHOD defining relationship to “Imaging - action (qualifier value)” (SCTID 360037004).

HAS_DESCRIPTION	Edge	An edge from a concept to its descriptions. A concept will have at least two descriptions -- a Fully Specified Name and a preferred term description.
HAS_ROLE_GROUP	Edge	An edge from a concept to the role groups which give semantic definitions for a concept.

Entity Relationships

The following diagram shows the structures in the database and the relationships between them.



Common terms and common elements used in the database structures

The term "SCTID" is used to indicate identifiers assigned by SNOMED CT to components of the SNOMED CT object model which contains Concepts, Descriptions and Relationships.

active and effectiveTime attributes

All SNOMED CT components have an associated "active" and "effectiveTime" attributes in any definition in an RF2 file. These attributes must be considered together and they indicate whether the definition was active or inactive at the specified date. The effectiveTime attribute is stored as a character string with the format YYYYMMDD (for example "20150901" for Sep 1, 2015). The active attribute uses value 1 to indicate an active definition and 0 to indicate an inactive definition.

id, id128 and sctid attributes

The id attribute stores the SCTID assigned by SNOMED CT for a concept or relationship component. Note that for description objects, the SNOMED CT assigned identifier is called id128 instead of id. The representation of id values are strings of decimal digits. The representation of id128 values has a hexadecimal digit representation (for example, 5298ada1-c7e5-5b51-be91-b11b46c6c9a2)

Components such as descriptions exist to provide information about other components such as a concept. Consider a description object in the database. The SNOMED CT identifier assigned to the description will be stored in the id attribute. The SNOMED CT identifier of the concept which it is defining is stored in the sctid attribute. A RoleGroup object is associated with a concept and has a sctid attribute which indicates the associated concept.

moduleId attribute values

These values indicate the source of the definition for a SNOMED CT component.

SCTID	FSN
900000000000207008	SNOMED CT core module (core metadata concept)
731000124108	US National Library of Medicine maintained module (core metadata concept)

Note that other moduleId values exist for all entities with localized SNOMED CT content.

history attribute

A string which encoding the historical modifications to the attributes associated with a SNOMED CT component. If the definition of the component has never been modified then the string is empty, otherwise it documents the initial attributes for the component and the history of changes.

nodetype attribute

These values are generated by the SNOMED_G application and are for internal usage.

Concept Nodes

The concepts defined in a Snomed CT release (in the sct2_Concept<suffix> file) are stored in nodes labeled as “ObjectConcept” which contain the following attributes:

Attribute name	Description	Source
		If blank – from Concept file, with the same field name as attribute.
active	Activation state of definition.	
effectiveTime	Effective date of definition.	
id	SCTID assigned to the concept.	
sctid	SCTID assigned to the concept (same as id in this case).	id
FSN	Fully Specified Name	FSN Term in Descriptions file, with matching sctid.
moduleId	Source of the definition (see table).	
nodetype	“concept”	Application generated.
definitionStatusId	Indicator of primitive concept versus fully-defined (see table).	
History	Component history.	Application generated.

definitionStatusId attributes

SCTID	FSN
900000000000074008	Necessary but not sufficient concept definition status (core metadata concept) NOTE: aka “primitive concept”
900000000000073002	Sufficiently defined concept definition status (core metadata concept) NOTE: aka “fully defined concept”

The following visualizes concept node 186791009 (Plasmodium falciparum malaria with cerebral complications), along with some of its relationships.

Here are its attributes in the US 2015 0901 release of SNOMED CT:

- **id:**186791009
- **sctid:**186791009
- **FSN:**Plasmodium falciparum malaria with cerebral complications (disorder)
- **moduleId:** 900000000000207008
- **effectiveTime:** 20020731
- **nodetype:** concept
- **definitionStatusId:** 900000000000073002
- **active:** 1
- **history:** [{"active": "1", "effectiveTime": "20020131", "id": "186791009", "definitionStatusId": "900000000000074008", "moduleId": "900000000000207008"}, {"definitionStatusId": "900000000000073002", "effectiveTime": "20020731"}]

Description Nodes

The descriptions for the SNOMED CT concepts are stored in nodes labeled as “Description”, and contain the following attributes (strings):

Attribute name	Description	Source
		If blank – from Description file, with the same field name as attribute.
nodetype	“description”	Application generated.
active	Activation state of definition.	
effectiveTime	Effective date of definition	
sctid	SCTID of the concept being described.	
id128bit	SCTID assigned to the description.	
moduleId	SNOMED CT module id	
term	Descriptive text describing the associated concept.	
languageCode	Language used in the text of the term, for example “en” for English.	
descriptionType	“FSN” for fully-specified named, “Preferred” for preferred term or “Synonym”	
refsetId	The SCTID of concept describing the “reference set” providing the description.	
caseSignificanceId	The SCTID of the concept describing case sensitivity of concept (see table).	
acceptabilityId	Indicates the acceptability of the term in the language or dialect specified by that Reference Set. Values include "preferred" and "acceptable" (see table).	
typeId	SCTID of concept which defines if the term is a synonym or the FSN.	
History	Component history.	Application generated.

caseSignificanceId attribute values

SCTID	FSN
900000000000020002	Only initial character case (core metadata concept)
900000000000017005	Entire term case sensitive (core metadata concept)
900000000000448009	Entire term case insensitive (core metadata concept)

typeId attribute values

SCTID	FSN
900000000000013009	Synonym (core metadata concept)
900000000000003001	Fully specified name (core metadata concept)

acceptabilityId attribute values

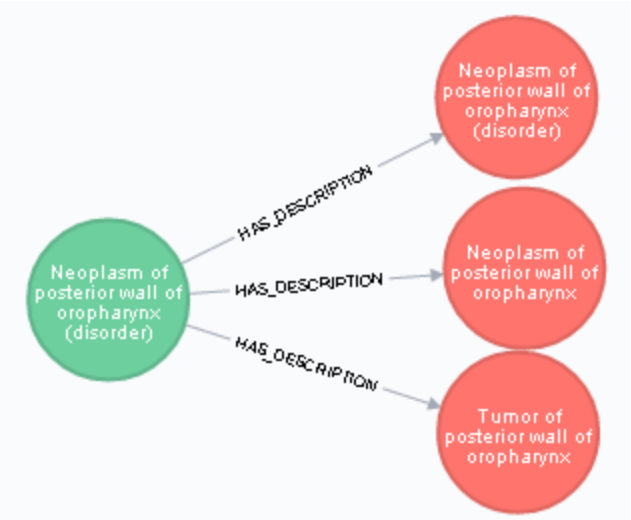
SCTID	FSN
900000000000548007	Preferred (foundation metadata concept)
900000000000549004	Acceptable (foundation metadata concept)

HAS_DESCRIPTION edges

An edge exists from each concept to each of its descriptions. There are minimally two Description nodes associated with a concept for the FSN and Preferred term definitions. Thus there are minimally two HAS_DESCRIPTION edges associated with each concept node. These edges are not defined in the Relationships file in the RF2 distribution and do not have associated SCTID values, as they are not strictly SNOMED CT components in themselves.

These edges have no attributes. They simply have a HAS_DESCRIPTION label and provide an edge from a concept to one of its descriptions.

The following is a visualization of the relationships between concept 126816002 (Neoplasm of the posterior wall of the oropharynx) and its descriptions within the database. The green node is the ObjectConcept and the orange nodes are the associated Description nodes, in this case -- FSN, Preferred Term and Synonym descriptions from top to bottom.



ISA relationships

ISA relationships documents the subtype relationships between SNOMED CT concepts. For example, the concept “Patella excision” (SCTID 125675003) is a subtype of “Operative procedure on knee ” (SCTID 6240004). The concepts involved are referred to as the source and destination concepts of the ISA relationship. This is documented by the ISA relationship which exists between those two concepts. In this case, “Operative procedure on knee” is a supertype of “Patella excision”.

Attribute name	Description	Source
		If blank – from Relationship file.
active	Activation state of definition.	
effectiveTime	Effective date of definition	
id	SCTID assigned to the ISA relationship.	
sctid	The SCTID of the source node of the ISA relationship.	sourceId
typeId	Relationship type SCTID, which is always 116680003 for ISA relationships.	
relationshipGroup	This is always zero for ISA relationships, which are not associated with role groups.	
moduleId	Source of the definition.	
characteristicTypeId	Defines whether the ISA Relationship forms part of the definition of the source concept (see table).	
History	Component history.	Application generated.

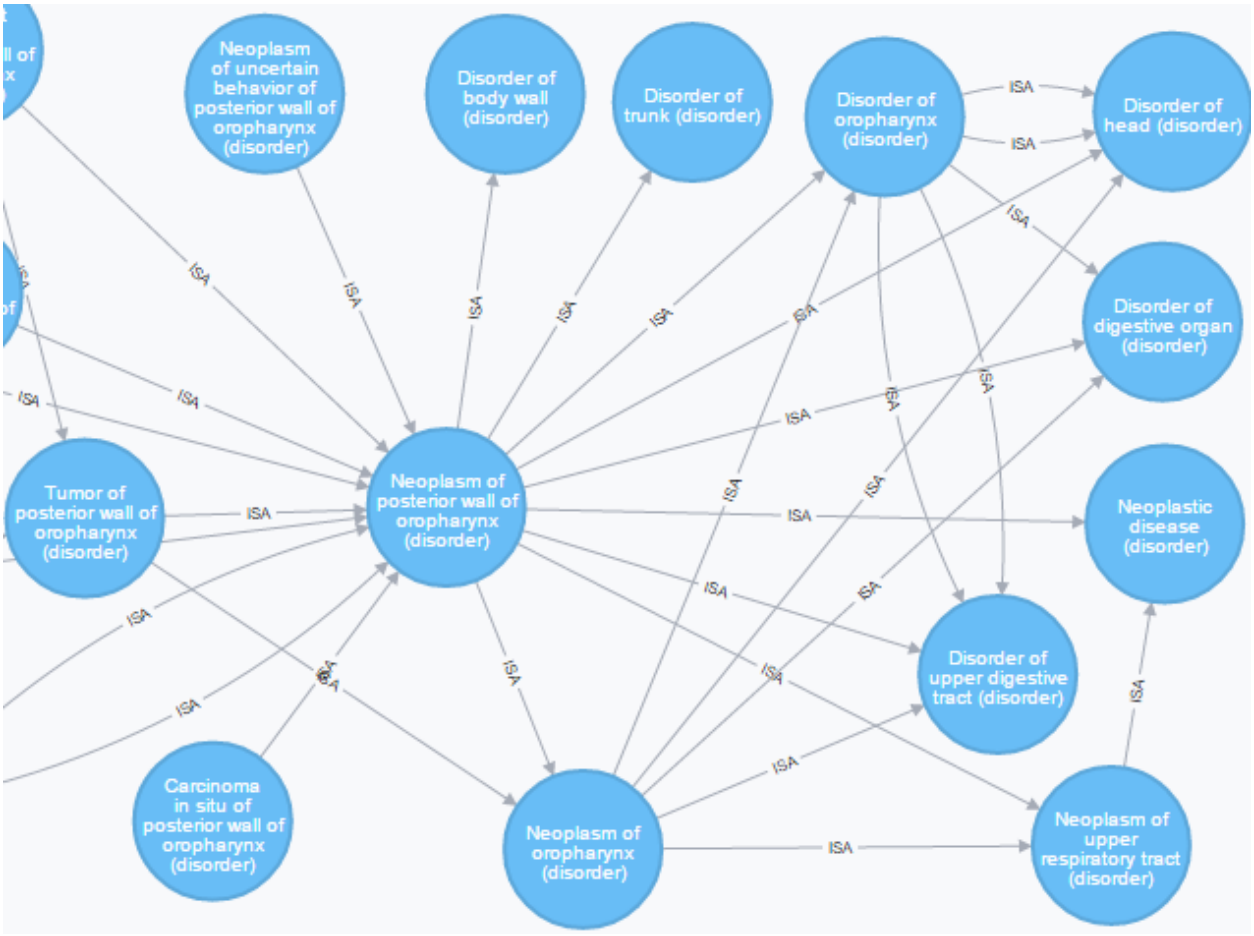
typeId attribute values

SCTID	FSN
116680003	Is a (attribute)

characteristicTypeId attribute values

SCTID	FSN
900000000000011006	Inferred relationship (core metadata concept)

The following visualizes some of direct incoming and outgoing ISA relationships associated with the concept 126816002 (Neoplasm of the posterior wall of the oropharynx), which gives an idea of the level of complexity of the ISA relationship definitions within SNOMED CT.



Defining Relationships

These relationships yield part of the semantic definition for a SNOMED CT concept. The semantics for a concept includes definitions inherited through ISA relationships, as well as the explicit content definition provided by the defining relationships.

The defining relationship edges are tied to role groups, where each role group is treated as an atomic semantic unit. The semantics derived from the defining relationships for a concept is the union of semantics from all of the role groups for a concept.

These relationships are edges between RoleGroup nodes and ObjectConcept nodes. Each defines part of a role group. There are over 1,000 defining relationship types. For example, a concept for medical procedure will commonly have a role group which has a METHOD defining relationship. For example, the procedure "CT of petrous bones (procedure)" (SCTID 241522007) has a role group with a METHOD defining relationship to "Imaging - action (qualifier value)" (SCTID 360037004).

Attribute name	Description	Source
		If blank – from Relationship file.
active	Activation state of definition.	
effectiveTime	Effective date of definition	
id	SCTID assigned to the ISA relationship.	
sctid	The SCTID of the source node of the ISA relationship.	sourceId
typeId	Relationship type SCTID, describing the nature of the relationships (see table).	
relationshipGroup	The role group number that the defining relationships is associated with.	
moduleId	Source of the definition.	
characteristicTypeId	Defines whether the ISA Relationship forms part of the definition of the source concept (see table).	
History	Component history.	Application generated.

characteristicTypeId attribute values

SCTID	FSN
900000000000011006	Inferred relationship (core metadata concept)
900000000000227009	Additional relationship (core metadata concept)

typeId attribute values

SCTID	FSN
260507000	Access (attribute)
370127007	Access instrument (attribute)
255234002	After (attribute)

260669005	Approach (attribute)
363715002	Associated etiologic finding (attribute)
246090004	Associated finding (attribute)
116683001	Associated function (attribute)
116676008	Associated morphology (attribute)
363589002	Associated procedure (attribute)
47429007	Associated with (attribute)
246075003	Causative agent (attribute)
263502005	Clinical course (attribute)
263535000	Communication with wound (attribute)
246093002	Component (attribute)
260908002	Course (attribute)
363699004	Direct device (attribute)
363700003	Direct morphology (attribute)
363701004	Direct substance (attribute)
42752001	Due to (attribute)
246456000	Episodicity (attribute)
260858005	Extent (attribute)
408729009	Finding context (attribute)
419066007	Finding informer (attribute)
418775008	Finding method (attribute)
363698007	Finding site (attribute)
127489000	Has active ingredient (attribute)
363705008	Has definitional manifestation (attribute)
411116001	Has dose form (attribute)
363702006	Has focus (attribute)
363703001	Has intent (attribute)
363713009	Has interpretation (attribute)
116678009	Has measured component (attribute)
116686009	Has specimen (attribute)
363710007	Indirect device (attribute)
363709002	Indirect morphology (attribute)
309824003	Instrumentation (attribute)
363714003	Interprets (attribute)
272741003	Laterality (attribute)
246267002	Location (attribute)
370129005	Measurement method (attribute)
367346004	Measures (attribute)
260686004	Method (attribute)
246454002	Occurrence (attribute)
246100006	Onset (attribute)
123005000	Part of (attribute)
370135005	Pathological process (attribute)
308489006	Pathological process (qualifier value)
260870009	Priority (attribute)
408730004	Procedure context (attribute)

405815000	Procedure device (attribute)
405816004	Procedure morphology (attribute)
363704007	Procedure site (attribute)
405813007	Procedure site - Direct (attribute)
405814001	Procedure site - Indirect (attribute)
370130000	Property (attribute)
370131001	Recipient category (attribute)
246513007	Revision status (attribute)
410675002	Route of administration (attribute)
370132008	Scale type (attribute)
246112005	Severity (attribute)
118171006	Specimen procedure (attribute)
118170007	Specimen source identity (attribute)
118168003	Specimen source morphology (attribute)
118169006	Specimen source topography (attribute)
370133003	Specimen substance (attribute)
258214002	Stage (attribute)
131195008	Subject of information (attribute)
408732007	Subject relationship context (attribute)
424876005	Surgical approach (attribute)
408731000	Temporal context (attribute)
363708005	Temporally follows (attribute)
370134009	Time aspect (attribute)
261583007	Using (attribute)
425391005	Using access device (attribute)
424226004	Using device (attribute)
424244007	Using energy (attribute)
424361007	Using substance (attribute)

Rolegroup Nodes

The role groups associated with a SNOMED CT concept are determined from the Relationships file. These are not official SNOMED CT components with their own SCTID values.

Each defining relationship has a relationshipGroup property which indicates the number of the role group that the relationship is assigned to. There can be multiple RoleGroup nodes associated with an ObjectConcept. A single RoleGroup is treated as an atomic semantic unit. The semantics of a SNOMED CT concept includes the union of semantics of all associated role groups, along with the semantics of concepts reachable by ISA relationships.

Attribute name	Description	Source
		If blank – from Description file, with the same field name as attribute.
nodetype	"rolegroup"	Application generated.
rolegroup	The role group number. This matches the relationshipGroup attribute from the defining relationships associated with this role group.	
sctid	SCTID of associated concept.	

HAS_ROLE_GROUP Relationships

An edge exists from each concept to each of its role groups. A concept may not have any role groups in which case it will not have any of these edges.

These edges have no attributes. They simply have a HAS_ROLE_GROUP label and provide an edge from a concept to one of its role groups.

The following visualizes the the relationships between concept 87821002 (Structure of apex of patella" and its role groups and defining relationships. In this case the relationships are straight-forward as there is a single defining relationship (PART_OF) associated with a single role group.