Medical Informatics in a United and Healthy Europe
K.-P. Adlassnig et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-044-5-223

Relationship Groups in SNOMED CT

Ronald CORNET [a,1], Stefan SCHULZ [b]
[a] *Department of Medical Informatics,*
Academic Medical Center – University of Amsterdam, The Netherlands
[b] *Institute of Medical Biometry and Medical Informatics,*
University Medical Center, Freiburg, Germany

Abstract. Relationship groups are a construct which is particular for the representation of concepts in SNOMED CT. In this paper, the July 2008 version of SNOMED CT is analyzed to determine the usage of relationship groups. Relationship groups are used with 36 out of 65 relations, playing a role in 28% of all concepts. Relationship groups are used in the concept types: "procedure" (including "regime/therapy"), "finding" (including "disorder"), "situation with explicit context" and "specimen". Examples are used to extract the purposes for using role groups, after which alternative approaches are discussed. The results indicate that relationship groups can be replaced by a small number of types of relationships such as "has sign" or "has subprocess".

Keywords. SNOMED CT, ontology, knowledge representation, description logic

1. Introduction

SNOMED CT is the inheritor of a dynasty of medical nomenclatures and coding systems, most notably SNOMED RT and Clinical Terms V3 [1]. The representation underlying SNOMED has been provided with formal rigor with the introduction of SNOMED RT. In 2002, relationship grouping was introduced as an extension to the description logic underlying SNOMED RT and SNOMED CT [2, 3]. This extension was motivated by the way in which concepts are modeled in SNOMED. In the first release of SNOMED CT, there were over 17,000 disorder concepts that are related to more than one morphology and more than one site, and there are more than 13,000 procedure concepts with more than one procedure-site and more than one method [3].

Recently, relationship groups have been discussed and alternatives have been proposed [4, 5]. In this paper we will analyze the current use of relationship groups in SNOMED CT, and determine for what kind of concepts their use is currently required. Based on this analysis, alternatives are discussed.

2. Relationship Groups

In this section a shortened version of the explanation given in [3] is presented. In SNOMED CT, relationships between concepts are specified to formally define

[1] Corresponding Author: Ronald Cornet, PhD, Department of Medical Informatics, Academic Medical Center, Room J1B-115, P.O.Box 22700, 1100 DE Amsterdam, The Netherlands; Phone: +31 (0)20 566 5188; E-mail: R.Cornet@amc.uva.nl.

concepts. Disease concepts are typically defined by a finding site and an associated morphology, such as:

PULMONARY_VALVE_STENOSIS : *FINDING_SITE* PULMONARY_VALVE,
ASSOCIATED_MORPHOLOGY STENOSIS

However, concepts can involve more than one finding site and more than one associated morphology. In these cases, definitions must be explicit regarding what morphology is related to which site. A frequently used example for this is the tetralogy of Fallot, which is a syndrome involving: right ventricle hypertrophy, aorta overriding, pulmonary valve stenosis, and incomplete closure of the interventricular septum.

When this is formally modeled, it must be prevented that it is interpreted as, for example, stenosis of the thoracic aorta.

Using relationship groups requires additional rules when performing reasoning (e.g., classification) on SNOMED CT. These rules are based on the interpretation that relationship groups can be seen as anonymous properties. Using this interpretation, the definition of pulmonary valve stenosis becomes:

disease: {associated_morphology=stenosis, finding_site= pulmonary_valve_structure}

or, using the description logic notation:

PULMONARY_VALVE_STENOSIS ≡ DISEASE ⊓
∃RG.(∃ *FINDING_SITE* PULMONARY_VALVE_STRUCTURE ⊓
∃ *ASSOCIATED_MORPHOLOGY* STENOSIS)

A consequence of this way of modeling is that tetralogy of Fallot will be classified as a subclass of pulmonary valve stenosis. This is said to be advantageous, but this claim will be addressed in the discussion.

3. Materials and Methods

3.1. SNOMED CT

The July 2008 release of SNOMED CT was used. The core of this release contains three text files. In this release, 383,230 concepts are defined, for which 1,134,773 English-language preferred and synonymous descriptions are provided. The concepts are defined using a total of 1,380,227 relationships.

The SNOMED CT relational source file contains the following information for each relationship:

- RelationshipID: a unique identifier for the relationship;
- ConceptID1: ID of the source concept of the relationship;
- RelationshipType: ID representing the type of relationship between the concepts;
- ConceptID2: ID of the target concept of the relationship;
- CharacteristicType: specifies whether a relationship is defining or a qualifier;
- Refinability: specifies whether the target concept can be refined (e.g., in data entry);
- RelationshipGroup: the group within the source concept that this relationship belongs to.

3.2. Methods and Analyses

The text files of the July 2008 release of SNOMED CT are imported into an MS Access database. Queries are created to perform analyses on the use of relationship groups, as well as analysis of concepts involved, in order to gain insight into the type of concepts and relationships for which relationship groups are used.

4. Results

4.1. Usage Analysis

SNOMED CT contains 315,550 active (i.e., current or limited) concepts. Of these, 87,345 (i.e., 28%) are defined using relationship groups.

Relationship groups only contain defining attribute relationships (i.e., no relationships that are qualifiers [6], and no *IS_A* relationships). Thirty six (out of 65) types of relationships are used in relationship groups, of which *FINDING SITE*, *ASSOCIATED MORPHOLOGY* and *METHOD* are the ones most used in relationship groups.

The number of relationship groups in a concept ranges from 0 for 81,914 concepts to 7 for 4 concepts, and 10 for the concept RASTELLI OPERATION IN REPAIR OF TRANSPOSITION OF GREAT VESSELS (ConceptID 18932005).

Relationships with group number 0 are considered non-grouped relationships. Analysis turns out that there are concepts with ungrouped relationships only, grouped relationships only, and with mixed grouped and ungrouped relationships.

Furthermore, the group number is not necessarily consecutive, and concepts are found which the lowest group number is 3, e.g., ABDOMINAL AORTOGRAPHY WITH BILATERAL ILIOFEMORAL ARTERIES WITH SERIALOGRAPHY (ConceptID 50964001).

Table 1 shows which types of concepts are defined with multiple relationship groups. The typology is based on the extension of the fully specified name (FSN) of the concepts, which generally identifies the top hierarchy. However, two FSNs in Table 1 represent subtypes; Regime/Therapy is a subtype of Procedure, and Disorder is a subtype of Finding.

4.2. Reconstruction of Generic Use Cases for Relationship Groups

Table 1 shows that relationship groups are used for six types of concepts. For each of these types a use case will be addressed.

"Procedure" concepts introduce relationship groups in order to separate various

Table 1. The type of concepts in which relationship groups are used. For each type, the number of concepts having relationship groups is given, the total number of concepts of that type, and the resulting percentage. Only concepts that have status "active" are included

Type	# concepts with relationship group	Total # concepts	% concepts with relationship group
Procedure	31,112	46,881	66%
Disorder	36,062	63,749	57%
Finding	6,010	32,638	18%
Situation	2,575	3,472	74%
Specimen	753	1,170	64%
Regime/Therapy	323	2,823	11%

subprocess of the procedure. In case of the above-mentioned Rastelli operation, this is a combination of anastomosis, bypass, graft, and repair and reconstruction of various sites.

"Regime/therapy" concepts provide groups of method/site combinations. In some of these concepts, the method is the same for all groups. For example, frowning exercises has one group (method: training ; site: structure of corrugator supercilii muscle) and one group (method: training ; site: structure of precerus muscle). Other concepts combine different methods, e.g., mobilization of the spine has group (method: manipulation ; site: regional spinal joint structure) and group (method: mobilization ; site: joint structure).

"Finding" concepts have groups consisting of morphology/site combinations, and interprets/has interpretation combinations. For example, venous stasis is defined with relationship groups (has interpretation: abnormal; interprets: vascular flow) and (has interpretation: decreased; interprets: vascular flow)

"Disease" concepts, like findings, generally consist of morphology/site combinations.

"Situation" concepts generally have relationship groups consisting of the explicit representation of context: associated finding, finding context, temporal context and subject relationship context. For example, "father smokes" has these respective values for the relationships: smoker, known present, current or specified, father.

"Specimen" concepts have relationship groups consisting of specimen source topography, specimen procedure, and specimen substance. For example, the concept "specimen from left kidney obtained by radical nephrectomy and adrenalectomy" is defined using two groups of relationships with respective values (left kidney structure, radical nephrectomy, body tissue material) and (structure of left adrenal gland, adrenalectomy, body tissue material).

5. Discussion

The initial analysis of the use of relationship groups in SNOMED CT illustrates that these groups can have different interpretations. For procedures, groups describe subprocesses or steps within the procedure [4]. For regimes/therapies, groups either distinguish various elements of a therapy, or indicate actions undertaken in order to result in a given therapeutic effect, e.g., manipulation of a regional spinal joint structure for the purpose of mobilization of a joint structure. For findings, groups address various interpretations (e.g., low, abnormal). Diseases consist of groups which generally reflect various symptomatic characterizations. For situations, the groups address various aspects or views on the situation described. For specimens, the groups describe the "ingredients".

The motivation for using (anonymous) groups of relationships rather than explicit relations such as *HAS_PART*, *HAS_SUBPROCESS*, *HAS_SIGN*, or *HAS_INGREDIENT* was that the latter would restrict reasoning capabilities. For example, as mentioned in the introduction, using relationship groups enables that tetralogy of Fallot is classified as a subclass of pulmonary valve stenosis. Consequently, when querying patients with pulmonary valve stenosis, patients with tetralogy of Fallot will also be found.

However, it can be argued that it is a case of IsA-overloading [7], and that it is more ontologically correct to represent a tetralogy of Fallot as consisting of a (rather than being a) pulmonary valve stenosis.

In the same vein, it is arguable that a pentalogy of Fallot is a subclass of tetralogy of Fallot.

Thus, when reconsidering relationship groups as various types of constituents of a concept, the *IS_A* link between TETRALOGY OF FALLOT and PULMONARY VALVE STENOSIS would be replaced by, e.g, a *HAS_SIGN* link. This would also give the user more flexibility in querying SNOMED CT encoded data. A query for pulmonary valve stenosis would then selectively retrieve the cases with an isolated stenosis, that is not part of a more complex clinical picture. To extend the query to retrieve also the latter cases the query should be expanded to $\exists$ *HAS_SIGN*. PULMONARY_VALVE_STENOSIS. Alternative, role chaining could be used to specify that *HAS_DIAGNOSIS* "traverses over" the *HAS_SIGN* relationship: *HAS_DIAGNOSIS* $\circ$ *HAS_SIGN* $\sqsubseteq$ *HAS_DIAGNOSIS*. As such role inclusion axioms are already allowed in the logic underlying SNOMED CT, this logic can remain unchanged.

It must be realized that there are many similar inferences to be made, which can and should not be solved by subsumption, but by implication. For example, a patient who underwent CLOSED REDUCTION OF FRACTURE OF FEMUR, will have suffered from a FRACTURE OF FEMUR. Likewise, a patient developing SEPSIS *DUE TO/AFTER* PNEUMONIA, will have (had) pneumonia, but this does not imply that the SEPSIS *IS_A* PNEUMONIA.

Further research is needed to gain more insight into the consequences this representation will have in practice, mainly for the purposes of querying patient data.

6. Conclusion

The use of relationship groups has increased significantly since they were first introduced into SNOMED RT. The analysis described in this paper shows that their use is restricted to only 6 types of concepts. Furthermore it is shown that the semantics of relationship groups can be made explicit by using relationship such as *HAS_PART* or *HAS_SIGN*. IsA-overloading is prevented in this way, but still the required reasoning capabilities can be maintained by these alternative approaches. The practical impact of reduced IsA-overloading needs to be further analyzed.

References

[1] Cornet, R., de Keizer, N. (2008) Forty years of SNOMED: A literature review. *BMC Medical Informatics and Decision Making* 8(Suppl 1):S2.

[2] Spackman, K.A. (2001) Normal forms for description logic expressions of clinical concepts in SNOMED RT. *AMIA Annual Symposium Proceedings 2001*, 627–631.

[3] Spackman, K.A. et al. (2002) Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED. *AMIA Symposium Proceedings 2002*, 712–716.

[4] Schulz, S. et al. (2006) The semantics of procedures and diseases in SNOMED CT. *Methods of Information in Medicine* 45(4):354–358.

[5] Schulz, S., Marko, K., Suntisrivaraporn, B. (2008) Formal representation of complex SNOMED CT expressions. *BMC Medical Informatics and Decision Making* 8(Suppl 1):S9.

[6] Cornet, R. (2009) Definitions and qualifiers in SNOMED CT. *Methods of Information in Medicine* 48(2):178–183.

[7] Ingenerf, J., Linder, R. (2006) Ontological principles applied to biomedical vocabularies. *European Federation for Medical Informatics Special Topic Conference 2006*, Timisoara, http://www.imbi.uni-freiburg.de/medinf/fctc-2006/docs/in_final.pdf.