# Generating SNOMED CT Subsets from Clinical Glossaries: An Exploration Using Clinical Guidelines

Carlos Rodríguez-Solano, Jesús Cáceres, and Miguel-Ángel Sicilia

Information Engineering Research Unit,
Computer Science Dept., University of Alcalá,
Ctra. Barcelona km. 33.6 – 28871 Alcalá de Henares (Madrid), Spain
{carlos.solano,jesus.caceres,msicilia}@uah.es

**Abstract.** The large SNOMED CT (SCT) terminology has gained adoption in the last years. However, its practical application for coding clinical information is hampered by its complexity and size. The mechanism of subsets allows for creating clusters of SNOMED CT terms that cover a particular application or clinical domain. These subsets are usually defined following some sort of consensual expert-driven process that is effort-intensive. The automated generation of subsets from clinical document corpora have been proposed elsewhere, but they still require a collection of documents that is representative for the targeted domain. This paper describes an experiment in using clinical guidelines' glossaries as a seed terminology for automatically generating subsets by traversing SNOMED relationships. Quantitative analysis reveals that traversing patterns need to be limited, and expert assessments point out that the approach may be viable at least for bootstrapping the process of elaborating the subsets.

**Keywords:** SNOMED CT, subsets, clinical guidelines, glossaries.

## 1 Introduction

In the last few years there has been a growing body of literature about the use of SNOMED CT [1] (Systematized Nomenclature of Medicine Clinical Terms) as an standard reference terminology aimed at achieving interoperability between clinical systems that can be implemented and used in different clinical settings. The sheer size of SNOMED CT is a significant issue in developing, using and maintaining it. So, extracting meaningful fragments from this terminology, is a key issue for using it. Yet, there are few detailed encoding instructions showing how this can be done and the issues involved.

Several techniques for extracting fragments of ontologies have been developed [2, 3]. Most of these techniques rely on employing various heuristics for determining which classes (concepts) are relevant and which are not. The algorithm implemented in the PROMPT-FACTOR tool [4] is one example; given a domain vocabulary and ontology, it retrieves an initial subset, and then, the vocabulary is expanded with the other atomic concepts or roles of the subset. Another example is the algorithm [5] which was used for segmentation of the medical ontology GALEN [6]. In [7] a user is

M.M. Cruz-Cunha et al. (Eds.): CENTERIS 2011, Part III, CCIS 221, pp. 117–127, 2011.

allowed to state what relations are to be considered in the procedure of making the subset. For refining the results to their own requirements the method allows the user to run the method with different configurations.

However this paper describes a heuristic method used for developing SNOMED CT subsets derived from clinical terms collected from clinical guidelines'(CG) glossaries. As illustration, we have used two specific-domain examples: first, the extraction of subsets including information relevant to the recognition and initial management of ovarian cancer, and then, subsets in the context of Parkinson's disease diagnosis and management in primary and secondary care.

The rest of this paper is structured as follows: in Section 2, background on the material to be used in this paper (clinical guidelines and SNOMED CT) is provided. The methods supporting our encoding process is given in Section 3; details about the steps involved in these methods are provided in Sections 3.1-3.4. In the fourth section quantitative results obtained by the encode method are described. We conclude with some remarks about future work.

## 2 Background

Clinical guidelines are recommendations based on the best available evidence on the appropriate treatment and care of people with specific diseases and conditions. They provide guidance and set quality standards for improving people's health and prevent and treat ill health.

Guidelines can provide recommendations for the treatment and care of people by health professionals, and also can be used to develop standards to assess the clinical practice of individual health professionals [8].

Clinical guidelines review a number of clinical questions which focus on areas of uncertainty or where there is a wide variation in clinical practice. These clinical questions were chosen using a consultative process involving patient groups, representatives from relevant professional organizations and the pharmaceutical industry. Therefore, the health professionals' knowledge included in clinical guidelines, is very representative about the particular domain covered, both in content and vocabulary used to express it. Clinical guidelines usually include glossaries which provide a list of specific clinical and medical terms which have been referred to in them.

SNOMED CT is a comprehensive, multilingual clinical reference terminology that allows healthcare providers to record clinical encounters accurately and unambiguously. It can be used to code, retrieve, and analyze clinical data. In January 2011 it contained over 390,000 concepts, 1,160,000 English descriptions and 1.40 million relationships. It consists of three core building blocks:

- **Concepts:** each *concept* represents a single clinical meaning
- **Descriptions:** each concept *description* is a term (a phrase used to name a concept) or name assigned to a SNOMED CT concept; any concept may have any number of descriptions
- **Relationships:** each *relationship* represents a logical relationship between two concepts

Numeric codes (ConceptID, DescriptionID and RelationshipID ) identify every instance of the three core building blocks. Unlike previous coding schemes, SNOMED CT is not a code-dependant hierarchy; instead, it relies on a large number of explicitly defined relationships. Moreover, multiple SNOMED CT concepts can be joined together to create post-coordinated expressions that allow users to record complex clinical conditions.

SNOMED CT is both a coding scheme, identifying concepts and terms, and a multidimensional classification. The content coverage of SNOMED CT is organized into 19 hierarchies including: clinical finding, procedure, observable entity, body structure, organism, substance, pharmaceutical/biological product, specimen, special concept, physical object, physical force, event, environment or geographical locations, social context, situation with explicit context, staging and scales, linkage concept, qualifier value and record artifact.

# 3 Methods

In this paper we have used the NICE (National Institute for Health and Clinical Excellence) clinical guideline CG122 (Full Guideline – April 2011)[1] which reviews a number of clinical questions that involve the detection, diagnosis and initial management of ovarian cancer, and, clinical guideline CG35 (Full Guideline – July 2011)[2] for Parkinson's disease diagnosis and management in primary and secondary care. Also, for this research, the January 31, 2011 International Release version of SNOMED CT was used.

The encoding process involves the following steps:

- SCT data preparation.
- Extract the clinical guidelines' glossary- this is the *input data set.*
- Encode the data items included in the input data set - this is an initial SNOMED subset named *encoded data set (EDS).*
- Apply "heuristic rules" to expand an input SNOMED subset with other SCT concepts and export the concept codes – in this step is generated a new *SNOMED CT subset*. This step can be iterated.

An overview of this method is shown in Figure 1.

## 3.1 SCT Data Preparation

SCT is distributed as text files. Hence this distribution data needs to be restored to a framework to identify SCT's elementary data structures. To manipulate the data through the computational framework is necessary a fast and consistent way. Therefore, the raw data, consisting of the three tables; concepts, descriptions and relationships, were placed into the relational database management system, MySQL.

[1] http://guidance.nice.org.uk/CG122

[2] http://guidance.nice.org.uk/CG35

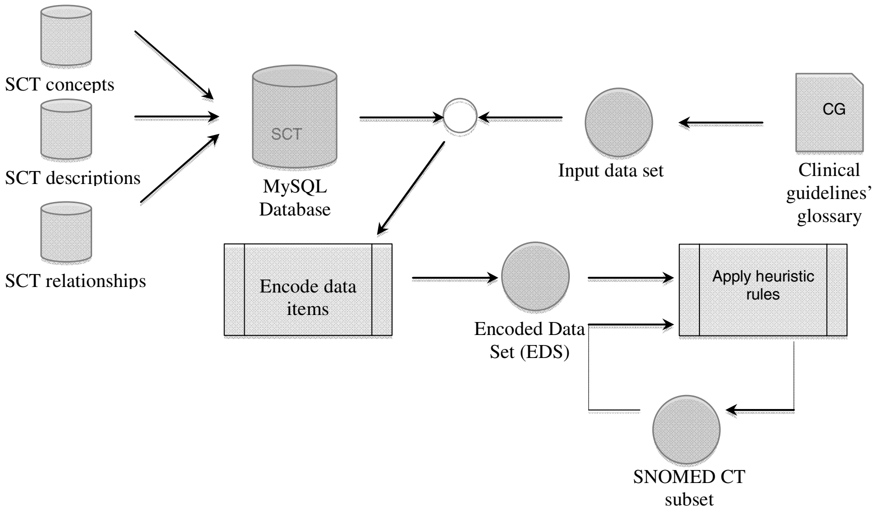

**Fig. 1.** An overview of the method for automatically generating SNOMED subsets from clinical glossaries

### 3.2 Extract the Clinical Guideline's Glossary

The input data set extracted from the clinical guidelines' glossary for detection, diagnosis and initial management of ovarian cancer (hereinafter denoted as "Input_CG122") contained 126 items; as cites, chemotherapy and gastro-splenic ligament are three data items examples contained in Input_CG122.The data set extracted from clinical guidelines'glossary for Parkinson's disease diagnosis (hereinafter denoted as "Input_CG35") contained 98 elements; in this case, sialorrhoea and dyskinesia are two glossary items examples.

### 3.3 Encode the Data Items Included in the Input Data Set

Lexical string matching is our method of locating the SNOMED CT concepts to identify all the SCT candidates. The matching algorithm takes each data item from the input data set and attempts to find each SCT Description in the description table of the MySQL database which matches.

Using the SNOMED CT fields *DescriptionStatus* and *ConceptStatus*, it checks that both the description and the concept are in active use.The algorithm returns results by exact match. Exact matches occur when all words are found in the SNOMED CT description and are in the same order as the data item. The SNOMED CT description must not contain additional words. The *encoded data set*( to be hereinafter denoted as "EDS")contains all the SCT concepts associated with each description matching with a data item; it is the initial SCT subset.

### 3.4 Apply “Heuristic Rules” to Expand an Input SNOMED CT Subset

This phase of the method is intended to extract relevant parts of SNOMED CT which are likely to be related to the clinic domain under consideration. In this sense, we have chose some “heuristic rules” which establish certain patterns to be applied for traversing SNOMED relationships; therefore an input SCT subset is expanded meaningfully according to experimental criteria.

SNOMED CT provides a rich set of inter-relationships between concepts, which are at the heart of it. Each relationship is defined as an object-attribute-value triple. The object, identified by a concept identifier (ConceptId1), is the source concept, the one that has the relationship. The attributes establish the type of the relationship (RelationshipType), and is also a SNOMED CT concept. The value is the target (ConceptId2).There are four categories of relationships: defining, qualifying, historical and additional. There are two types of defining relationships:

- *Super - Subtype relationships*, also known as *IS_A relationships* or *Parent-Child relationships*

These relationships define specific concepts as children of more general concepts (parents) providing the main semantic hierarchy that relates concepts to one another. The Super - Subtype relationship concept has 116680003 as conceptID and FSN(FullySpecifiedName) |is a|. A given concept (Concept_X) may have subtype children (concepts with a subtype relationship referring to Concept_X ) or supertype parent (concepts referred to by a subtype relationship from Concept_X).

- *Attribute relationships*

These relationships contribute to establish defining characteristics about a concept (ConceptId1); the RelationshipType indicates the nature of the defining attribute, and ConceptId2 represents the value of that attribute.

Therefore, SNOMED CT Concept Model is a multiaxial, hierarchical classification system; its graph structure , with concepts and relationships represented as nodes and arcs, has a main axis, which contributes the hierarchical type based aspect of a concept definition, and another concepts associated with concept’s characteristics.

The heuristic rules chosen in our study, were induced by SNOMED CT axes mentioned above, and bearing in mind that, when they are applied, cause an expansion of *the input SCT subset* just along one dimension, using heritage criteria, or, through two dimensions, applying the latter together with concept’s defining characteristics.

Next, the defined heuristic rules, denoted as “*HIp*” and “*HAj*” are provided; the application of each heuristic rule gives rise to a SNOMED CT subset denoted as “*SCT_HIp*” or “*SCT_HAj*”.

- *HIp*(p=1, 2, 3) : For each concept (node) contained in EDS, generate all level jconcepts (vertices), $p \geq j \geq 1$, through relationship 116680003|is a|;the application of this rule allow us to obtain SNOMED subsets with increasing granularities, ie, concepts representing increasingly specific levels of detail are added to it; in this way, a new SNOMED subset denoted by SCT_HIp is obtained.

- *HAj* (j=1, 2, 3) : For each concept Concept_X contained in *SCT_HIp* , with j = p, generate all concepts (attributes' values), using relationships with RelationshipType ≠116680003, referring to Concept_X ; SCT_HIp is expanded and SCT_HAj is obtained.

When rules are applied, duplicate concepts are eliminated from the output SCT subset.

**For instance**

The word *Chemotherapy* is an element included in Input_CG122, the set which contains the items extracted from the clinical guidelines' glossary for detection, diagnosis and initial management of ovarian cancer. This input data item matches with the SNOMED CT concept 367336001|chemotherapy|; it belongs to the semantic category Procedure, one of the 19 Top-level hierarchies, through which SNOMED CT content is organized. Therefore, concept 367336001|chemotherapy| is an EDS member.

Then, if the heuristic rule named HI2 is chosen to be applied, the encoded data set (EDS), which is the input SNOMED subset, is expanded; all level j, 2 ≥ j ≥ 1, concepts are added to EDS. In Figure 2 is shown some descendants of concept 367336001|chemotherapy|, ie concepts which are elements of SCT_HI2 :

- 266719004|oral chemotherapy|, 265762008|subcutaneous chemotherapy| and 265760000|intravenous chemotherapy| are level 1 descendants (children)
- 51534007|oral chemotherapy for malignant neoplasm| is a level2 descendant

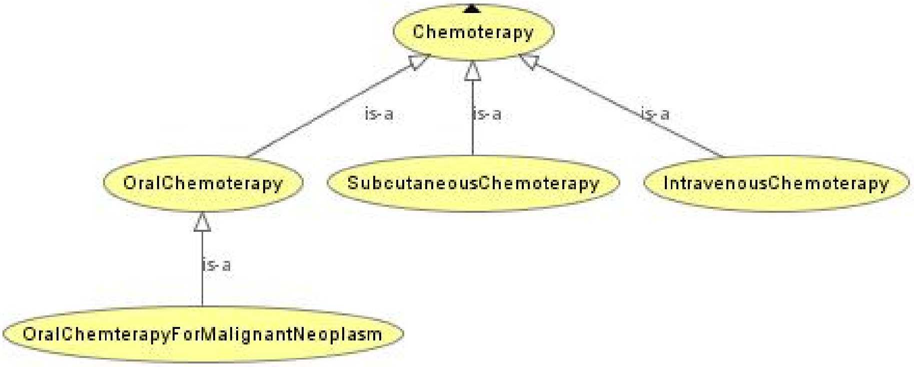

**Fig. 2.** Some descendants of concept 367336001|chemotherapy| (each concept node belongs to SCT_HI2)

Once the Snomed subset SCT_HI2 has been obtained, let us proceed to apply HA2 heuristic rule, considering SCT_HI2 as the input subset ; thus, SNOMED subset named SCT_HA2 is generated.

Fig. 3 shows the new graph (fragment) obtained:

- 51534007|oral chemotherapy for malignant neoplasm (procedure)|:
  363703001|has intent (attribute)| = 262202000|therapeutic intent (qualifier value)|

The *attribute* (has intent) specifies the intent of the *procedure* (oral chemotherapy for malignant neoplasm) and the *value* (therapeutic intent) indicates the nature of the procedure.

- 51534007|oral chemotherapy for malignant neoplasm|: 363701004|direct substance (attribute)| = 410942007|drug or medicament (substance)|

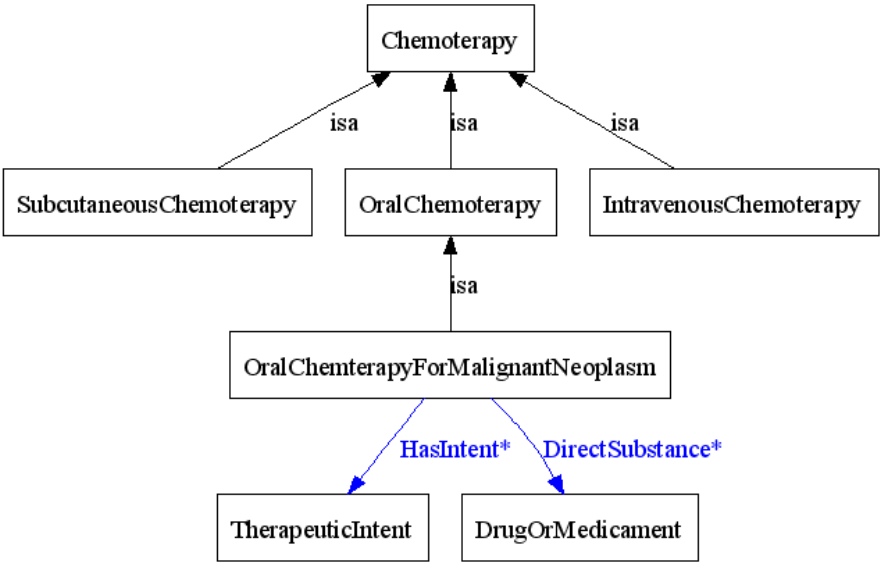

**Fig. 3.** Graph (fragment) whose vertices represent concepts included in SCT_HA2

## 4 Results and Discussion

The following section describes quantitative results obtained by the encoding method described above, both in the case of clinical guideline CG122 (ovarian cancer) as in the case of CG35 (Parkinson's disease).

The input set Input_CG122 ( glossary items) contained 126 elements and Input_CG35 contained 98 items; in each case, Table 1 indicates that 71 (≈ 53%) and 20 (≈ 20%) glossary data items were matched with SNOMED CT concepts, respectively.Examples of unencoded glossary terms are available in Table 2.Therefore, the number of SNOMED CT concepts contained in the encoded data set, EDS, is 71 and 20 for each of the case studies.

**Table 1.** A summary of the input data set (glossary items) and matching terms

| | **CG122** | **CG35** |
|---|---|---|
| No. of clinical terms collected from clinical guidelines' glossary | 126 | 98 |
| No. of exact matches | 71 | 20 |

**Table 2.** Sample glossary items that could not be encoded with SNOMED CT

| Glossary item(CG122) | Glossary item(CG35) |
|---|---|
| Cytology | Lee Silverman Voice Treatment (LSVT) |
| Gastro-splenic ligament | Motor fluctuations |
| Percutaneous core biopsy | Cochrane Review |

Based on the initial SNOMED CT subset EDS, traversing Is_a (HIp heuristic rules) relationships recursively, three new subsets were generated: SCT_HI1, SCT_HI2 and SCT_HI3. They are interrelated by the set inclusion relation between sets; their cardinalities are shown in Table 3.

The SNOMED CT subset covering 53% (20%) of codable glossary items, plus an expansion of descendants with increasing order of granularity intended to extract relevant parts of SNOMED, contains 1,203 (100), 7,199 (142) and 15,117 (153) SNOMED concepts. The SCT release (January 2011) has 293,670 active concepts; therefore, these expansions' cardinalities correspond to 0,4 % (0,03%), 2,4% (0,03%) and 5,14% (0,05%) of the content, respectively. Therefore, the coverage of these subsets uses a maximum of 5,14 (0,05) percent of SCT.

**Table 3.** Encoding summary statistics (EDS : Encoded data set)

| Input SNOMED CT subset | No. of concepts in input subset | | Heuristic rule applied | Output SNOMED CT subset | No. of concepts in output subset | |
|---|---|---|---|---|---|---|
| | CG122 | CG35 | | | CG122 | CG35 |
| EDS | 71 | 20 | HI1 | SCT_HI1 | 1,203 | 100 |
| EDS | 71 | 20 | HI2 | SCT_HI2 | 7,199 | 142 |
| EDS | 71 | 20 | HI3 | SCT_HI3 | 15,117 | 153 |
| SCT_HI1 | 1,203 | 100 | HA1 | SCT_HA1 | 1,622 | 141 |
| SCT_HI2 | 7,199 | 142 | HA2 | SCT_HA2 | 8,607 | 184 |
| SCT_HI3 | 15,117 | 153 | HA3 | SCT_HA3 | 17,628 | 197 |

The *set difference* of sets A and B, and, the *intersection* of A and B, written A \ B and A ∩ B respectively, where A and B are SNOMED CT subsets, were computed for quantitatively assessing the degree of the semantic expansion [9] between two subsets, which were obtained using the method detailed above. The results are shown in Table 4. Some obvious results, or results easily inferred from table 3, have been omitted.

Hereinafter, we denote the cardinality of the set A as #(A). The following quantitative analysis take in account data appearing in the second column of Table 4.

The SNOMED CT subset SCT_HA3, which is the one with maximum cardinality, comprises 17, 628 SNOMED concepts (table 3), and uses 6 percent of SCT. This subset was generated expanding SCT_HI3 subset by the addition of 2511 new concepts to it, representing attributes' values, among them,1408 were attributes' values referring to concepts included in SCT_HI2.Also, SCT_HI3 increased on 7918 concepts with respect to SCT_HI2. Hence, the degree of the semantic expansion measured in terms of concepts, between SCT_HA3 and SCT_HI2, is 10,429 concepts.

**Table 4.** Set difference and Intersection of SNOMED CT subsets

| Subsets Operations | No. of concepts/CG122 | No. of concepts/CG35 |
|---|---|---|
| SCT_HI1 \ EDS | 1,132 | 80 |
| SCT_HI2 \ EDS | 7,128 | 122 |
| SCT_HI3 \ EDS | 15,046 | 133 |
| SCT_HA1 \ SCT_HI1 | 419 | 41 |
| SCT_HA1 \ SCT_HI2 | 389 | 41 |
| SCT_HA1 \ SCT_HI3 | 358 | 41 |
| SCT_HI2 \ SCT_HA1 | 5966 | 42 |
| SCT_HA2 \ SCT_HI1 | 7,406 | 84 |
| SCT_HA2 \ SCT_HI2 | 1,408 | 42 |
| SCT_HA2 \ SCT_HI3 | 1,291 | 42 |
| SCT_HA3 \ SCT_HI1 | 16,427 | 97 |
| SCT_HA3 \ SCT_HI2 | 10,429 | 55 |
| SCT_HA3 \ SCT_HI3 | 2,511 | 44 |
| SCT_HI1 ∩ SCT_HA1 | 1,203 | 100 |
| SCT_HI1 ∩ SCT_HA2 | 1,201 | 100 |
| SCT_HI1 ∩ SCT_HA3 | 1,201 | 100 |
| SCT_HI2 ∩ SCT_HA1 | 1,233 | 100 |
| SCT_HI2 ∩ SCT_HA2 | 7,199 | 142 |
| SCT_HI2 ∩ SCT_HA3 | 7,199 | 142 |
| SCT_HI3 ∩ SCT_HA1 | 1,264 | 100 |
| SCT_HI3 ∩ SCT_HA2 | 7,316 | 142 |
| SCT_HI3 ∩ SCT_HA3 | 15,117 | 153 |

Taken in account that in SNOMED CT some classes are defined in extension (i.e. via a list of their subclasses) rather than in intension (i.e. via a list of characteristics) [10], we considered appropriated to express the semantic expansion by two kinds of "semantic components" [11] : 1. Hierarchical-component : 7918 concepts (i.e. component weight ≈0,76) and 2. Roles-component: 2511 concepts (i.e. component weight ≈ 0,24). In a similar way, we can state the semantic expansion between SCT_HA3 and SCT_HA2 : 1. Hierarchical-component : 7918 concepts (i.e. component weight ≈ 0,87) and 2. Roles-component: 1103 concepts (i.e. component weight ≈ 0,13).

The relative growth rates associated with each expansion are decreasing :(one dimension) SCT_HI1&SCT_HI2 ≈ 4,98 and SCT_HI2&SCT_HI3≈ 1,09; (two dimensions) SCT_HI1&SCT_HA1≈ 0,34, SCT_HI2&SCT_HA2 ≈ 0,19 and SCT_HI3&SCT_HA3 ≈ 0,16. In contrast, the absolute changes are increasing : SCT_HI1&SCT_HI2 = 5996 concepts and SCT_HI2&SCT_HI3 = 7918 concepts ; SCT_HI1&SCT_HA1 =419 concepts, SCT_HI2&SCT_HA2 =1,408 concepts and SCT_HI3&SCT_HA3 = 2,511 concepts. These results mean that although increments produced in each expansion are elements of an increasing sequence, these growths represent a decreasing proportion respect to the cardinality of the corresponding input subset.

Now, we consider data which appear in the third column of Table 4.

For the second case (clinical guideline CG35), the SNOMED CT subset SCT_HA3, which is the one with maximum cardinality, comprises 197 SNOMED concepts (table 3), and uses 0,06 percent of SCT.

From #(SCT_HI3 ∩ SCT_HA2) = 142 and #(SCT_HA2 \ SCT_HI3) = 42, together with

#(SCT_HA2) = 184 (Table 3), we can infer that all target concepts (attribute's value) belonging to SCT_HA2 has level less than or equal to two in a different hierarchy than that of the source concept. The same happens with SCT_HA1. Instead, for the other case, using data appearing in the second column of Table 4, #(SCT_HI3 ∩ SCT_HA2) = 7,316 and#(SCT_HA2 \ SCT_HI3) = 1,291, together with

#(SCT_HA2) = 8,607 (Table 3), imply that 117 target concepts belonging to SCT_HA2 has level three in a different hierarchy than that of the source concept .

## 5 Conclusions

This paper has explored the use of clinical guidelines' glossaries as a seed terminology for automatically generating subsets covering a clinical domain by traversing SNOMED relationships. Quantitative analysis reveals two main facts: (1) the application of heuristic rules needs to be limited and (2) as the rules operate in different "dimensions" the choice of the kind of rule to be applied has to be carefully balanced.

Although our primary concern in this pilot was to make quantitative analysis related to the intrinsic characteristics of the method, expecting to obtain a brief assess about the benefits of it linked to semantic boundary delimitation , two clinicians reviewed a small number of concepts included in each of the generated subsets, evaluating their relevance to the chosen domain. In this sense, the conclusions were not definitive, but they point out that the approach may be viable as starting point in the process of elaborating subsets.

The data shown in the third column of Table 4 shows that the order of magnitude is smaller compared with data appearing in second column. This suggests that the nature of the disease being considered in the chosen clinical guideline has important implications when generating SNOMED subsets.

Future work will deal with an extensive evaluation of different approaches for generating subsets under experimental settings.

Also, an important point of care to be considered, is the use of natural language processing techniques, which may allow this method to be refined in the encoding phase by generating a more accurate encoded data set from the items in the glossaries.

Although still in a preliminary stage the work has established that it is feasible to construct useful initial subsets for SCT using clinical guidelines' glossaries, enabling the development of a methodology for deriving SNOMED subsets.

**Acknowledgements.** This work has been supported by the project "Métodos y Técnicas para la Integración Semántica de Información Clínica mediante Arquetipos y Terminologías (METISAT)", code CCG10-UAH/TIC-5922, funded by the Community of Madrid and the University of Alcalá.

Likewise, this research work has been partially supported by projects PI09-90110 - PITES - from "Fondo de Investigación Sanitaria (FIS) Plan Nacional de I+D+i", and DGPY-1301/08, "Plataforma en Red para el Desarrollo de la Telemedicina en España" from FEDER.

## References

1. International Health Terminology Standards Development Organisation, http://www.ihtsdo.org/snomed-ct/
2. Cuenca Grau B., Horrocks I., Kazakov Y., Sattler U.: Extracting modules from ontologies: Theory and practice. Technical report, University of Manchester (2007), http://www.cs.man.ac.uk/bcg/Publications.html
3. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Just the Right Amount: Extracting modules from ontologies. In (IW3C2), WWW 2007, Banff, Alberta, Canada (May 8-12, 2007)
4. Noy, N., Musen, M.: The PROMPT suite: Interactive tools for ontology mapping and merging. Int. Journal of Human- Computer Studies 6(59) (2003)
5. Seidenberg, J., Rector, A.: Web ontology segmentation: Analysis, classification and use. In: Proc. WWW 2006 (2006)
6. Rector, A., Rogers, J.: Ontological issues in using a description logic to represent medical concepts: Experience from GALEN. In: Proc. of IMIA WG6 Workshop (1999)
7. Noy, N.F., Musen, M.A.: Specifying ontology views by traversal. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 713–725. Springer, Heidelberg (2004)
8. National Institute for Health and Clinical Excellence. About clinical guidelines, http://www.nice.org.uk/aboutnice/whatwedo/aboutclinicalguidelines/about_clinical_guidelines.jsp
9. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: A logical framework for modularity of ontologies. In: Proc. IJCAI 2007, pp. 298–304 (2007)
10. Bodenreider, O., Barry Smith, B., Anand Kumar, A., Anita Burgun, A.: Investigating Subsumption in SNOMED CT: An Exploration into Large Description Logic_Based Biomedical Terminologies. Artificial Intelligence in Medicine 39, 183–195, PMC2442845 (2007)
11. Cimino, J.J.: Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf. Med. 37, 394–403 (1998)