MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230928

International Patient Summary Terminology

Warren DEL-PINTO[a,1], Renate A. SCHMIDT[a], Yongsheng GAO[b], Ghadah ALGHAMDI[a,c], Alejandro Lopez OSORNIO[b] and Suzy ROY[b]
[a]*Department of Computer Science, University of Manchester, UK*
[b]*SNOMED International, UK*
[c]*Department of Computer Science, Dar Al-Hekma University, Kingdom of Saudi Arabia*
ORCiD ID: Warren Del-Pinto https://orcid.org/0000-0003-3307-9432, Renate A. Schmidt https://orcid.org/0000-0002-6673-3333, Yongsheng Gao https://orcid.org/0000-0002-3468-2930, Ghadah Alghamdi https://orcid.org/0000-0003-3836-1824

Abstract., SNOMED CT is a comprehensive medical ontology used in health care sectors across the world covering a wide range of concepts that support diversity at the point of healthcare. However, not all these concepts are needed for every use case; it is better to concentrate on those parts that apply to the particular application while preserving the meaning of relevant concepts. This paper considers the application of a novel subontology extraction method to create a new resource, called the IPS terminology, which functions as a standalone ontology with the same features as SNOMED CT, but is designed for cross-border patient care. The IPS terminology has been released for free use under an open license, with the intention of promoting interoperability of health information worldwide.

Keywords. Medical terminology, ontology, International Patient Summary, SNOMED CT, subontology generation

1. Introduction

SNOMED CT is a comprehensive clinical terminology containing a broad range of concepts that supports diversity at the point of healthcare in different countries across the world. As such, SNOMED CT is by necessity a large domain ontology; it contains over 350,000 clinical concepts and over 350,000 axioms that define the meaning of and relationships between these concepts, making use of the Web Ontology Language OWL2 to provide formal semantics for this purpose [1].

For many healthcare applications, it is not necessary to make use of the entire terminology. Instead, it is beneficial to focus on parts of the ontology that are relevant to a given domain. Therefore, substantial effort is made by clinical experts to curate subsets of SNOMED CT concepts, called reference sets (refsets). A prominent example of such initiatives is the International Patient Summary (IPS) [2]. While many refsets are curated to capture data in a specialist domain, the IPS is a minimal, non-exhaustive set of concepts that capture the basic, speciality-agnostic and condition-independent

[1] Corresponding Author: Warren Del-Pinto, email: warren.del-pinto@manchester.ac.uk

data of patients. A key aspect of this is that these concepts are readily usable by all clinicians for the purpose of unscheduled (cross-border) patient care.

While useful in narrowing the focus for tasks such as data entry and analysis, flat lists of concepts are not sufficient to take advantage of the semantics of SNOMED CT expressed in OWL2. The semantics enables tasks such as classification to capture stated and inferred relationships between clinical concepts and querying that enables identification of populations within data corresponding to expressive specifications. Therefore, it is better to create a standalone subontology of the source ontology that summarizes the semantics of the concepts. Modularization approaches [3-5] can compute subsets of the axioms within an ontology, called modules, given a set of input concepts. However, when applied to an ontology as large as SNOMED CT with a large refset such as the IPS, the resulting modules are often too large and contain information that is not required for the given clinical purpose.

This paper considers the use of a new approach [6] to producing extracts of SNOMED CT, called subontology extraction, that has been applied to produce a new resource: the IPS Terminology. The IPS Terminology can be used in applications requiring the hierarchical information and querying capabilities available for the full SNOMED CT ontology, but focused on the domain of cross-border patient care. In member countries, SNOMED CT is a free access terminology. The ability to extract a functional part of the larger SNOMED CT ontology was required in order to provide free access to this portion of the content to users in non-member countries.

2. Methods

When producing extracts of SNOMED CT for a given clinical use case, there are several important factors that should be taken into consideration. The extract should be concise; it should not contain large portions of the terminology that are not required to capture information specifically about the domain of interest. Instead, it should make available comprehensive definitions for the collection of concepts that are considered by domain experts to be essential to capturing information in their intended applications. These definitions should conform to the structural specifications as stated in the modeling standards [7,8], to aid in ensuring the consistency of modeling efforts, clarity of the definitions and integration with other SNOMED CT content. It is also important to preserve hierarchical links that exist between concepts in the computed extracts, so that they can be navigated as a standalone ontology.

To meet the needs of the SNOMED CT community, a new approach to producing standalone subontologies of SNOMED CT was developed in [6]. The approach takes as input a version the SNOMED CT, referred to as the source ontology, and a set of SNOMED CT concepts which are referred to as the focus concepts; these can be an existing refset or a new set of concepts. The approach produces subontologies that satisfy the following *three core conditions*: (i) Definitions and relationships of focus concepts in the source ontology are captured in authoring form satisfying SNOMED CT Editorial Guidelines. (ii) The subontology can contain supporting concepts and attributes besides focus concepts. (iii) The transitive closure/concept hierarchy over focus and supporting concepts in SNOMED CT must be reflected in the subontology.

The focus concepts are treated as the core of the domain of interest, and effectively determine the scope of the coverage provided by the resulting subontology. The approach only includes additional concepts outside the set of focus concepts if they are required to satisfy the conditions specified above. These additional concepts are called supporting concepts, and include the concepts used to provide definitions of the focus concepts. By satisfying these conditions, the aim is to produce extracts that emphasize conciseness while preserving the core definitions in the domain of interest, as well as the hierarchy between the concepts in the computed extract.

The separation of focus and supporting concepts in this way is a key difference between this approach and other commonly used modularization approaches [2-4]. Modularization aims to produce a subset of the source ontology, which includes all axioms for not only the concepts in the input signature but also any concepts that are introduced when computing the module. For subontology extraction in SNOMED CT, it is not necessary to provide full definitions for all of the supporting concepts, only those that are required to satisfy the three core conditions. As a result, the size of the resulting extracts usually differs between modularization approaches and subontology extraction as considered here. Given a set of focus concepts, subontologies tend to be smaller than the corresponding modules [6], making them more suited to the task of information capture in specific clinical applications.

An overview of the process of computing a subontology, given a source ontology in OWL format and a set of focus concepts from the ontology as input, is as follows:

1. **Compute Focus Concept Definitions.** These are given by the set of authoring form definitions (as axioms) [9] computed for the focus concepts and added to a new ontology, the subontology. This step adds new concepts outside the input set, supporting concepts, as part of the definitions of the focus concepts.
2. **Definition Expansion.** For each supporting concept introduced, and based on the above conditions, determine if the full authoring form definition of the supporting concept is required. If it is required, it is computed and added to the subontology. If additional supporting concepts are added as a result, these must be added to the set of supporting concepts to be checked during this step.
3. **Populate the Attributes.** It is then necessary to add the defining axioms for each of the attributes added to the subontology. To do this, a STAR module is computed [10] taking the set of attributes in the subontology as input.
4. **Add Grouper Concepts.** To support human navigation of the subontology, grouper concepts are added to organize the concepts in the subontology into related sub-hierarchies. By default, relevant top-level SNOMED CT categories are added, such as Clinical finding, Event, Procedure and so on.
5. **Complete the Concept Hierarchy.** The relationships, including ancestor/descendent relationships, are already captured for the focus concepts by steps 1–3 without editing the authoring form definitions. However, for the supporting concepts, it is necessary to complete the transitive closure, with respect to subsumption, to ensure that this information is complete in the subontology. To do this, subsumption relationships between concepts in the subontology are detected in the source ontology (using a reasoner) and, if missing from the subontology, the hierarchical relationships are added.
6. **Add Textual Descriptions.** Any "non-logical" information in the source ontology brought into the subontology, such as synonyms and textual definitions of concepts, are added to the subontology.

For the purposes of interoperability, it is also possible to compute the Release Format 2 (RF2) format [11] of the subontology from OWL2, which requires that the necessary normal form [7] is computed for each concept in the subontology. To ensure that all of the content available in the latest edition of SNOMED CT can be included in the extracted subontology, the approach supports the latest OWL language constructs used in the International Edition of SNOMED CT. These include general concept inclusions, attribute chains and transitivity and reflexivity of attributes.

3. Results

The subontology extraction approach described in the previous section was used to create an IPS Subontology, as part of the Internal Patient Summary initiative of SNOMED International. To create the IPS Subontology, the input was the current version of the International Edition of SNOMED CT (July 31st, 2022) in OWL format and the content of the IPS Freeset (refset), a list of 8,658 concepts, as the focus concepts. The subontology extraction process then produces a subontology containing 15,932 (focus and supporting) concepts that complete the SNOMED CT hierarchies, attributes and values of models, and some SNOMED CT metadata (Table 1).

Table 1. Example of concepts not present in the IPS Freeset, added to the IPS Terminology by the subontology extraction process (5,669 concepts in total), and the descendant concept (member of the IPS Freeset) that determined its inclusion.

Supporting concept added in the subontology	Freeset descendant concept (focus concept)
148911000119107 |Primary malignant neoplasm of abdomen (disorder)|	312104005 |Cholangiocarcinoma of biliary tract (disorder)|
609638001 |Operation on pelvic region of trunk (procedure)|	22523008 |Vasectomy (procedure)|
442571000124108 |Tree nut (substance)|	256350002 |Almond (substance)|

In subontologies the information is greatly compressed as not all the ancestors and descendants of the focus concepts are included (Table 2). The count of ancestors of concepts in the IPS Freeset in SNOMED CT is 7,725, whereas the same count in the resulting subontology is 1,976. Almost 5,000 concepts were not necessary to adequately represent the concept definitions and excluded for conciseness.

Table 2. Example ancestors of IPS Freeset concepts not included in the subontology for conciseness, as they are not required to represent the concept definition. In the "Almond" example, all ancestors were required.

Ancestors not included in the subontology	Freeset descendant concept (focus concept)
733355004 |Primary adenocarcinoma of digestive organ (disorder)|	312104005 |Cholangiocarcinoma of biliary tract (disorder)|
120013000 |Vas deferens excision (procedure)|	22523008 |Vasectomy (procedure)|
All ancestors were included	256350002 |Almond (substance)|

4. Discussion

This paper presents an approach to creating a SNOMED CT subontology to support the IPS project. The result, the IPS terminology, has been released as a free product with an open license, enabling interoperability of health information worldwide. The IPS terminology supports key features of SNOMED CT while focusing on the intended use case of the project. All concepts in the IPS freeset have equivalent definitions to those in the SNOMED CT terminology without loss in relevant semantics. ECL queries can be performed against the IPS terminology instead of SNOMED CT. ECL queries are often used for terminology binding specifications in information models such as FHIR. These same specifications can be applied to the IPS terminology. The only difference is that the number of concepts is restricted to the scope of the IPS terminology. For example, the ECL query << 105590001 |Substance (substance)| can be specified for the causative agents of allergies in FHIR, returning just over 500 relevant substances in the IPS terminology, whereas there are over 27,000 substances in the full SNOMED CT.

5. Conclusions

Future work will include making the subontology extraction tool easier to use and more accessible to end users, and applying code improvements and performance tuning.
Warren Del-Pinto was partially supported by UKRI/EPSRC grant EP/V047949/1.

References

[1] SNOMED International, SNOMED CT OWL Guide http://snomed.org/owl Accessed: 2022-11-29

[2] Kay S, Cangioli G, Nusbaum M. The international patient summary standard and the extensibility requirement. In: pHealth 2020: Proceedings of the 17th International Conference on Wearable Micro and Nano Technologies for Personalized Health; 2020 Sep 30; IOS Press; 2020. p. 54-62, doi: 10.3233/SHTI200615.

[3] Cuenca Grau B, Horrocks I, Kazakov Y, Sattler U. Modular reuse of ontologies: theory and practice. IEEE/ACM Trans. Audio, Speech, Language Process. 2008;31: 273-318, doi: 10.1613/jair.2375.

[4] Konev B, Lutz C, Walther D, Wolter F. Model-theoretic inseparability and modularity of description logic ontologies. Artif Intel. 2013;203: 66-103, doi: 10.1016/j.artint.2013.07.004.

[5] Rector AL. Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. In: K-CAP '03: Proceedings of the 2nd International Conference on Knowledge Capture; ACM; c2003. p. 121-128, doi: 10.1145/945645.945664.

[6] Del-Pinto W, Schmidt RA, Gao Y. Extracting subontologies from SNOMED CT. In: The Semantic Web: ESWC 2022 Satellite Events; 2022 May 29–June 2; Hersonissos, Crete, Greece, Cham: Springer International Publishing; 2022 Jul. p. 291-4, doi: 10.1007/978-3-031-11609-4_43.

[7] SNOMED International. Generating Necessary Normal Form Relationships from the OWL Refsets:https://confluence.ihtsdotools.org/display/DOCOWL/2.5.+Generating+Necessary+Normal+Form+Relationships+from+the+OWL+Refsets Accessed: 2022-11-29

[8] SNOMED International. SNOMED CT Concept Model: https://confluence.ihtsdotools.org/display/DOCSTART/6.+SNOMED+CT+Concept+Model Accessed: 2022-11-29

[9] Spackman KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. In: American Medical Informatics Association Annual Symposium. AMIA; c2001; p. 627-31.

[10] Sattler U, Schneider T, Zakharyaschev M. Which Kind of Module Should I Extract? In: Proceedings of the 22nd International DL Workshop; Oxford, UK: AMIA, vol 477; c2009; p. 627-31.

[11] SNOMED International. Release Schedule and File Formats: https://confluence.ihtsdotools.org/display/DOCSTART/13.+Release+Schedule+and+File+Formats Accessed: 2022-11-29