

Automated clinical pathway standardization using SNOMED CT- based semantic relatedness

Digital Health
Volume 8: 1–17
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076221089796
journals.sagepub.com/home/dhj



Ayman Alahmar¹ , Mohannad AlMousa¹ and Rachid Benlamri¹

Abstract

The increasing number of patients and heavy workload drive health care institutions to search for efficient and cost-effective methods to deliver optimal care. Clinical pathways are promising care plans that proved to be efficient in reducing costs and optimizing resource usage. However, most clinical pathways are circulated in paper-based formats. Clinical pathway computerization is an emerging research field that aims to integrate clinical pathways with health information systems. A key process in clinical pathway computerization is the standardization of clinical pathway terminology to comply with digital terminology systems. Since clinical pathways include sensitive medical terms, clinical pathway standardization is performed manually and is difficult to automate using machines. The objective of this research is to introduce automation to clinical pathway standardization. The proposed approach utilizes a semantic score-based algorithm that automates the search for SNOMED CT terms. The algorithm was implemented in a software system with a graphical user interface component that physicians can use to standardize clinical pathways by searching for and comparing relevant SNOMED CT retrieved automatically by the algorithm. The system has been tested and validated on SNOMED CT ontology. The experimental results show that the system reached a maximum search space reduction of 98.9% within any single iteration of the algorithm and an overall average of 71.3%. The system enables physicians to locate the proper terms precisely, quickly, and more efficiently. This is demonstrated using case studies, and the results show that human-guided automation is a promising methodology in the field of clinical pathway standardization and computerization.

Keywords

Automation in health care, clinical pathway, data analytics, SNOMED CT, ontology, health information system, public health care, semantic relatedness, semantic score

Submission date: 19 August 2021; Acceptance date: 9 March 2022

Introduction

Clinical pathway (CP) is a multidisciplinary, structured health care plan in which therapeutic and diagnostic medical interventions performed by doctors, nurses, and other staff members for a specific procedure or diagnosis are sequenced on a timeline.^{1–3} CPs can reduce physicians' mental effort and cognitive load to allow them to focus on thought-requiring, more complex health care activities.³ Therefore, CPs have the potential to improve patient outcomes and satisfaction. CPs also contribute to reducing the length of stay (LOS) in hospitals and controlling overall public health care costs.^{3,4}

Automation of modern health care systems covers many fields and necessitates the computerization of hospital

processes to streamline health care services, reduce paperwork, collect digital data for data analytics, and control costs^{5,6}. The CP was not an exception in this regard, as many studies reported the development of computerized CPs^{7–16}. By analyzing the literature, we found that the common theme in most studies is that the computerization process was directed mainly toward either connecting CPs with Electronic Medical Record (EMR) systems or

¹Department of Software Engineering, Lakehead University, Thunder Bay, Ontario, Canada

Corresponding author:

Ayman Alahmar, Assistant Professor, Department of Software Engineering, Lakehead University, Thunder Bay, Ontario, Canada P7B 5E1.
Email: aalahmar@lakeheadu.ca



computerizing CPs without standardizing their contents. This is viewed as a limitation because the focus of previous research was not on CP digitization; rather, the emphasis was mainly on EMRs or producing isolated CP systems that are difficult to integrate with all other digital health care systems. Table 1 is a comparison table that shows the limitations of CP computerization without standardization.

Therefore, one crucial area that remains to be investigated in CP automation is CP digitization, which can be facilitated by standardizing the clinical terms used in CPs. However, there is a gap today between the clinical terms used to define CP interventions and the clinical terms used in health information systems (HISs). On one hand, CPs use institutions' local informal terms, but on the other hand, HISs use and refer to standardized terminology terms such as SNOMED Clinical Terms (Systematized Nomenclature of Medicine – Clinical Terms, abbreviated

as SNOMED CT or SCT)¹⁷, which is the CP standardization terminology system adopted in this research. Non-standardized CPs are difficult to share across hospitals. In addition, non-standardized CP terms cannot be captured in HISs. Missing CP data due to this gap form a challenge to using data analytics in health care. The existing CP terminology gap can be traced to the following reasons:

1. CPs are developed internally in hospitals through internal meetings attended by local staff members.
2. CPs are commonly developed for internal use. Thus, the emphasis is on a language that is understandable by the institution's own health care members and administrators.
3. CPs are usually circulated in paper-based document forms, or tables and charts that are distributed to staff members and posted on poster boards, or on the doors of in-patients' hospital rooms.

The situation in many hospitals is that once CPs are developed and circulated internally, there is a future demand to computerize them.^{10,18} At that point, the terminology gap already exists which complicates the computerization and integration process. The terminology gap can be addressed by correctly standardizing CP terms so that CPs can be digitized and easily integrated with other HISs. In general, local CP terms that need to be standardized can be divided into three categories:

1. Category 1: Local CP terms that exist in SNOMED CT but do not accurately match the correct SNOMED CT terms (i.e. are not approved by domain experts as the correct SNOMED CT terms). For example, the SNOMED CT term "Intracranial hemorrhage (SCTID: 1386000)" was used incorrectly for the approved term "Subarachnoid intracranial hemorrhage (SCTID: 21454007)."
2. Category 2: Local CP terms that exist in other terminology systems (e.g. LOINC [19], ICD-10 [20], and ICD-10-PCS [21]) and mapping these terms to the adopted system is required (e.g. SNOMED CT terms in this study). For example, mapping the ICD-10-PCS term "Resection of Appendix, Percutaneous Endoscopic Approach (0DTJ4ZZ)" to the SNOMED CT term "Laparoscopic appendectomy (SCTID: 6025007)."
3. Category 3: Local CP terms that are nonexistent in any terminology system.

The objective of this research is to address the challenges in CP term standardization of Category 1 and Category 2 by exploiting semantic relations embedded in SNOMED CT. Standardization reduces missing CP data (i.e. CP data that are not captured in EMRs or other HISs), enhances the use of data analytics in health care, and reduces medical

Table 1. Limitations of CP computerization without standardization.

	Computerization without CP standardization	Computerization with CP standardization
CP Integration With HISs	Difficult to integrate due to mismatch between terms in CPs and terms used in HISs.	Easy to integrate since CPs share the same terminology standards with various HISs.
CP Update in EMRs	Connection is lost and rework is required to connect CP to EMR whenever the CP is updated.	Updated standardized terms integrate smoothly with counterpart EMR terms. No rework is required and the connection is not lost
CP Sharing Across Different Health Care Institutions	The terminology conflict creates ambiguity and makes CP sharing impossible.	CP standardization streamlines CP sharing without ambiguities.
CP Data Management and Reporting	Complex data management and reports are susceptible to errors due to conflicts.	Easier data management and standardized reports can be generated easily without errors.

CP: clinical pathway; EMR: electronic medical record; HISs: health information systems.

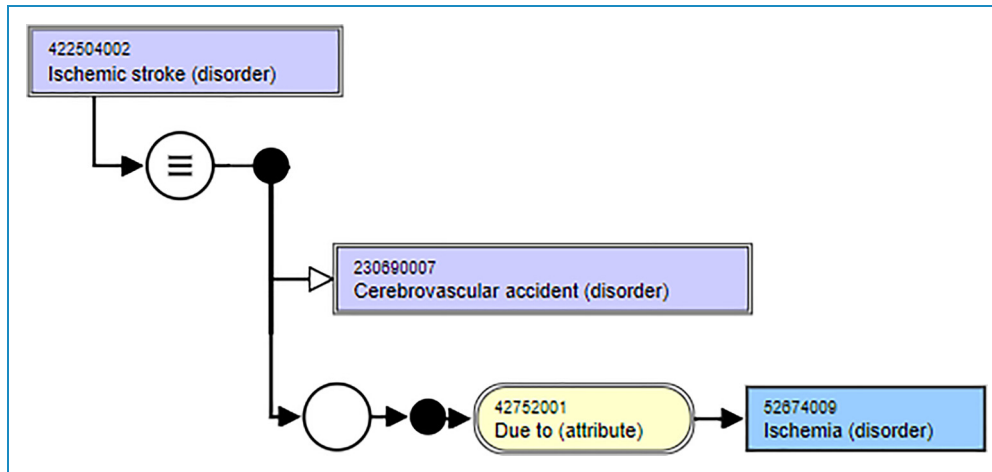


Figure 1. Part of the ischemic stroke concept hierarchy in SNOMED CT ontology.¹⁷

errors in hospitals. Addressing Category 3 local CP term standardization requires different AI techniques, which we will be considered in future work. Our domain experts in this research were physicians and nurses from the Regional Stroke Unit at the Thunder Bay Regional Health Sciences Centre in Thunder Bay, Ontario, Canada.

In Category 1, domain experts perform a tedious manual search in SNOMED CT hierarchy to locate the correct standardized terms. In Category 2, finding an initial similar term from SNOMED CT is straightforward, but locating an accurate SNOMED CT term based on the initial term is performed through a tedious manual search. A solution to this standardization challenge would be to automate the process of searching for accurate SNOMED CT standardized terms based on local/inaccurate initial terms. CP standardization is a challenging task to automate by machines due to patient safety considerations associated with the medical data stored in CPs. In this research, we hypothesize that the concepts of semantic relatedness, through the use of semantic relations in SNOMED CT ontology, can help humans in solving this standardization challenge by intelligently automating the tedious process of searching for and locating the most probable terminology corrections.^{22,23} The use of ontology semantic relations is justified by the fact that initial terms are similar in their meaning and semantic to the target terms (e.g. referring to the same body parts, similar types of medical conditions, etc.). The proposed approach in this research adopts SNOMED CT terminology system and achieves the balance between automatic machine intervention and domain experts control to ensure patient safety. Semantic relatedness is a form of measurement that quantitatively identifies the level of connectedness between two concepts based on all existing semantic relations.²³ On the other hand, semantic similarity is a metric that can be defined as a quantitative measure of likeness between terms based on their taxonomic distribution within a domain ontology.²²

Table 2. Top classes of SNOMED CT ontology.

Top classes of SNOMED CT Ontology	
Body Structure	Qualifier Value
Clinical Finding	Record Artifact
Environment or Geographical Location	Situation with Explicit Context
Event	SNOMED CT Model Component
Observable Entity	Social Context
Organism	Special Concept
Pharmaceutical/Biologic Product	Specimen
Physical Force	Staging and Scales
Physical Object	Substance
Procedure	

SNOMED CT: Systematized Nomenclature of Medicine - Clinical Terms.

Both semantic relatedness and semantic similarity are metrics over the terms; however, semantic relatedness includes any relation between the terms, while semantic similarity only includes “is-a” relations.^{22,23} For example, in SNOMED CT, “ischemic stroke”, whose SNOMED CT ID (SCTID) is 422504002, is similar to “cerebrovascular accident”, whose SCTID is 230690007 (there is an “is-a” relation between the terms in the ontology), but is also related to “ischemia”, whose SCTID is 52674009 (there is a “Due-to” relation between the terms in the ontology), as shown in Figure 1.¹⁷ Our method in this research includes

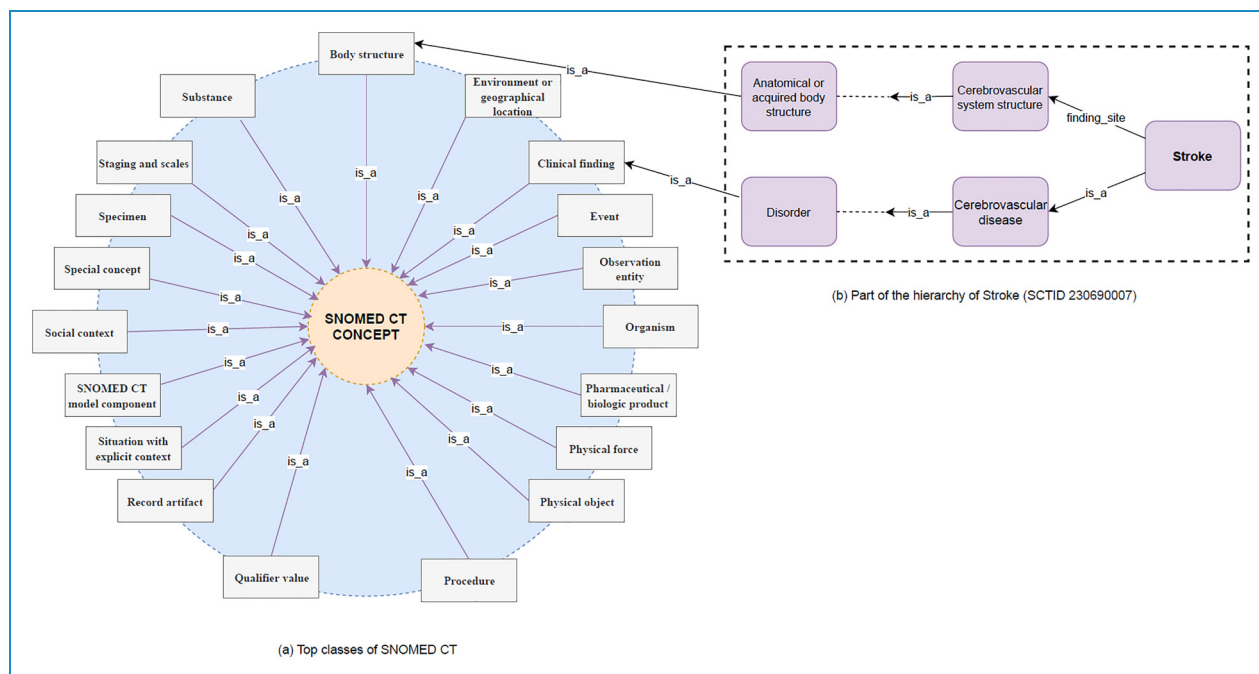


Figure 2. Top classes of Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) ontology.

Table 3. Example SNOMED CT relations and their definitions.

Relations used to define clinical finding concepts	Relations used to define procedure concepts
finding site: Specifies the body site affected by a condition	procedure site: Describes the body site acted on or affected by a procedure
due to: Relates a clinical finding directly to a cause such as another clinical finding or a procedure	method: Represents the action being performed to accomplish the procedure
finding method: Specifies the means by which a clinical finding was determined	procedure device: Describes the devices associated with a procedure
severity: Used to subclass a clinical finding concept according to its relative severity	using substance: Describes the substance used to execute the action of a procedure

SNOMED CT: Systematized Nomenclature of Medicine - Clinical Terms.

all semantic relations in SNOMED CT ontology (not restricted to is-a relations), and thus, it is a semantic relatedness approach. As highlighted in AlMousa et al. and Liu et al. [24,25], semantic relations play an intrinsic role in computing the semantic relatedness, thus a semantic score based on the various semantic relations embedded in SNOMED CT ontology is used in this research.

It is important to mention that since CPs are treatment plans of interventions and procedures performed on patients, caution should be exercised when utilizing automatic standardization of medical terms in CPs for safety reasons. Therefore, machine-based standardization methods should be considered as decision support methods rather than decision-making methods. The final decision-makers in medical CP term standardization are terminology-knowledgeable human domain experts. Therefore, in this paper, we present an approach where a list of terms is proposed by our algorithm; the final decision, however, is made by the domain experts. The proposed approach was validated by detailed case studies and by a dataset of 14 pairs of SNOMED CT terms (see the section on Experimental Results and Discussion and Table 6 for more details). The major contribution of this paper can be summarized as follows: The field of CP digitization and standardization is a new and emerging field with a few publications.^{26–28} To the best of our knowledge, this is the first study attempting to enhance the field with the automation of the CP term standardization process. Our semantic score approach is a holistic multi-relational approach that can explore all types of relations in the SNOMED CT ontology.²⁴

The rest of the paper is organized as follows. The next section addresses technical background and related work, and in particular, it covers SNOMED CT and CP Computerization and Standardization. Our method of CP standardization using SNOMED CT-based semantic relatedness is presented next and is followed by the experimental

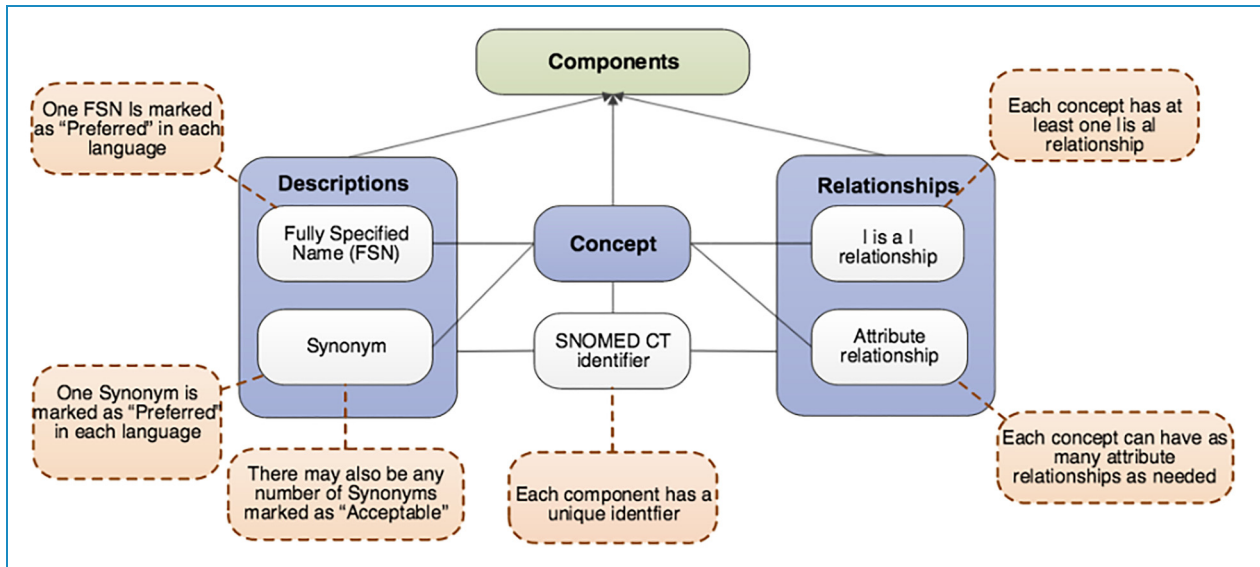


Figure 3. Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) logical model.¹⁷

Clinical Pathway	
Ischemic Stroke	
On Admission - Day 1	
Is swallowing screen completed?	Yes No - Complete swallowing screen
If patient failed swallowing screen, consult SLP	
If patient failed initial swallowing screen and SLP unavailable for assessment Repeat swallowing screen within 24 hours of admission	
If patient failed repeat swallowing screen: consult with physician regarding insertion of nasogastric tube/initiation of feeds as per enteral nutrition policies	
If on PO diet: Monitor and document tolerance to first meal	
Smoking Cessation	
Has patient used of any kind of tobacco in the last 6 months?	Yes No
If yes, discuss with patient and physician nicotine replacement options	
Patient and Family Education	
Provide patient and family with:	
- Your Stroke Journey: TOH Care Companion Booklet	
- Your Stroke Journey: A guide for people living with stroke	

Figure 4. Part of an ischemic stroke clinical pathway (CP) used at The Ottawa Hospital.

results and discussion. Finally, conclusions are drawn and future research work is presented in the last section.

Background and related work

SNOMED clinical terms

The Systematized Nomenclature of Medicine – Clinical Terms (officially abbreviated as SNOMED CT, and sometimes referred to as SCT) is a systematically organized computer processable collection of medical terms providing codes, terms, definitions, and synonyms that are used in reporting and medical documentation. SNOMED CT was

adopted in this work as the base of CP standardization because it is the most comprehensive health care clinical terminology in the world²⁹. The main purpose of SNOMED CT is to encode the meanings that are used in health care/health informatics and to support the effective clinical recording of data with the goal of improving patient care and health decision-making. SNOMED CT comprehensive coverage includes body structures, organisms, clinical findings, symptoms, diagnoses, procedures, and other etiologies, substances, pharmaceuticals, devices, and specimens.

Table 2 shows the names of the nineteen (19) top classes in the structure of the SNOMED CT ontology. Top classes

Table 4. Comparison with recent research.

Ref	Year	CP computerization model	CP term standardization	Terminology system used	Dataset used to validate standardization
[26]	2020	A generic framework for CP computerization and standardization supported with a CP-specific coding system to facilitate data analytics and machine learning in health care	Manual	SNOMED CT	N/A
[39]	2021	A computerized CP integrated with a standard medical report system for the treatment of inpatients with schizophrenia	Manual	ICD-10	N/A
[7]	2022	Integrating CP with an electronic health record system to improve adherence to COVID-19 hospital care guidelines	N/A (not mentioned in the paper)	N/A	N/A
This study	2022	An algorithm for automated term standardization to facilitate CP computerization and integration with HISs	Automated	SNOMED CT	A dataset of 14 pairs of SNOMED CT terms

CP: clinical pathway; SNOMED CT: Systematized Nomenclature of Medicine – Clinical Terms.

The screenshot displays a user interface for a SNOMED CT concept. The main header shows the concept name "Ischemic stroke (disorder)" with a star icon and a menu icon. Below this, the SCTID "422504002" is listed, followed by a breadcrumb trail "422504002 | Ischemic stroke (disorder) |". A list of synonyms is provided: "en Ischemic stroke", "en Ischaemic stroke", and "en Ischemic stroke (disorder)".

To the right, a box lists non-taxonomic relations:

- Finding site → Brain structure
- Associated morphology → Infarct
- Finding site → Cerebrovascular system structure
- Due to → Ischemia

Below these elements, a section titled "Children (10)" lists ten related concepts, each with a yellow circular icon and a blue minus sign:

- Cerebral ischemic stroke due to extracranial large artery atherosclerosis (disorder)
- Cerebral ischemic stroke due to intracranial large artery atherosclerosis (disorder)
- Cerebral ischemic stroke due to occlusion of extracranial large artery (disorder)
- Cerebral ischemic stroke due to small artery occlusion (disorder)
- Cerebral ischemic stroke due to stenosis of extracranial large artery (disorder)
- Ischemic stroke with coma (disorder)
- Ischemic stroke without coma (disorder)
- Ischemic stroke without residual deficits (disorder)
- Occlusive stroke (disorder)
- Perinatal arterial ischemic stroke (disorder)

Figure 5. Sample SNOMED CT (SCT) term with its non-taxonomic relations.

have is-a relations with the root class “SNOMED CT Concept,” as shown in Figure 2. Besides is-a relations, SNOMED CT concepts have many other relations/attributes such as associated-with, contained-in, due-to, finding-

site, has-ingredient, and is- about. Relations in SNOMED CT are organized based on their roles in the ontology (e.g. relations used to define clinical finding concepts, relations used to define procedure concepts, relations used to

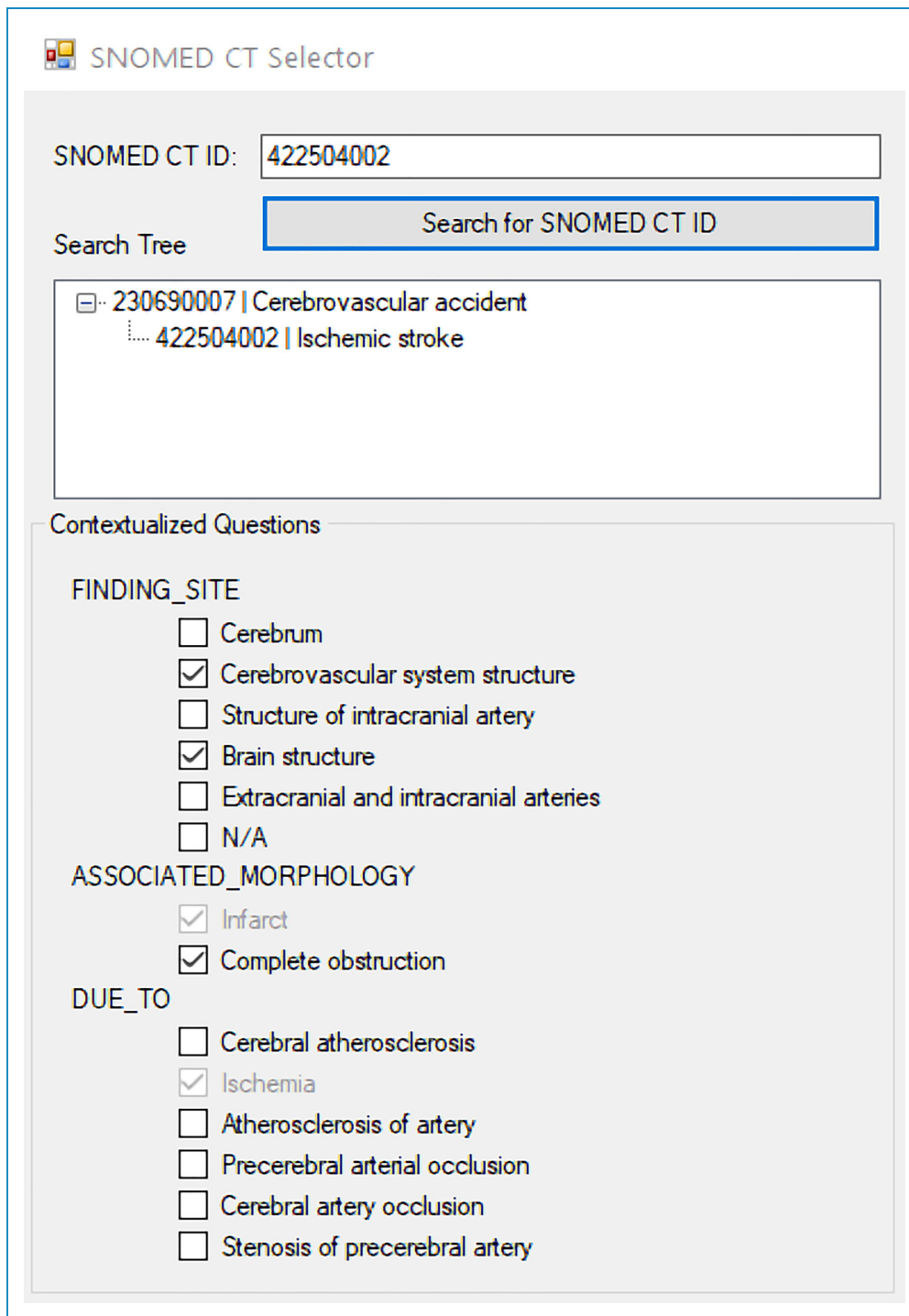


Figure 6. Contextualized relevancy questions based on the “Ischemic stroke” child SNOMED CT (SCT) terms.

define body structure concepts). Table 3 shows example relations and their definitions.

The January 31, 2021 release of the SNOMED CT International Edition included 350,000+ concepts that provide the core general terminology for electronic health records (EHRs). The SNOMED CT logical model (Figure 3) defines the way in which each type of

component and derivative is related and represented in SNOMED CT. The core component types are concepts, descriptions, and relationships. The logical model therefore specifies a structured representation of the concepts used to represent clinical meanings, the descriptions used to refer to these concepts, and the relationships between the SNOMED CT concepts.¹⁷

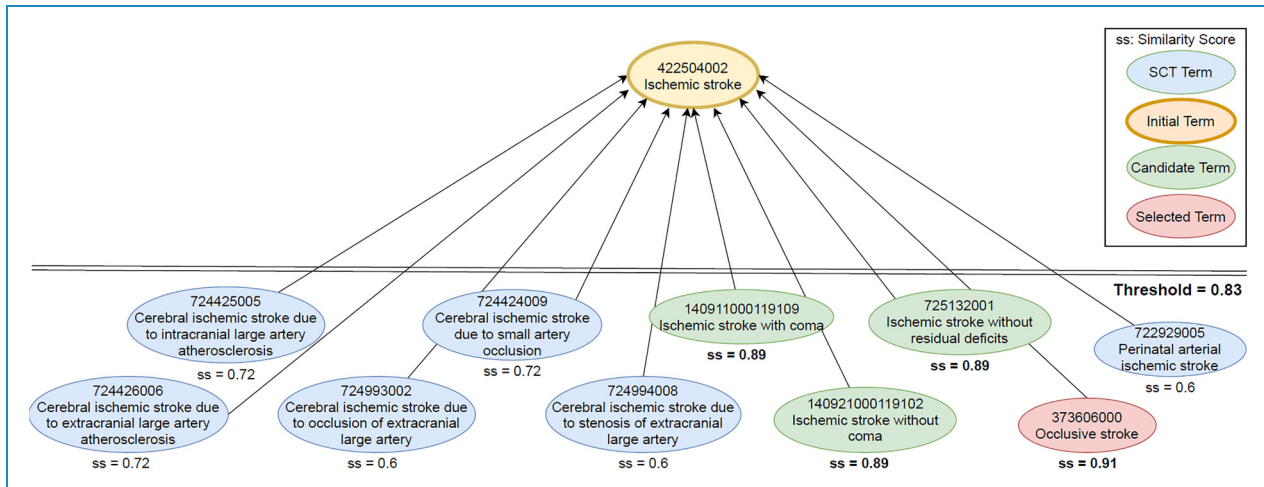


Figure 7. Children of the “ischemic stroke (SCTID: 422504002)” SNOMED CT (SCT) term and their similarity scores (*SimScore*).

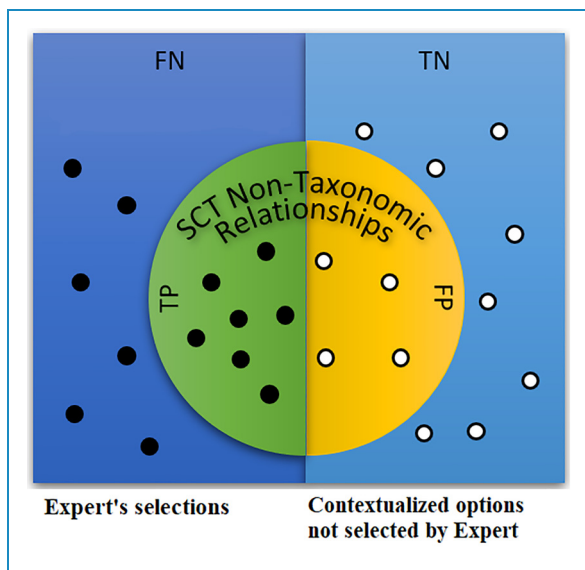


Figure 8. Illustration of the use of F1 as a semantic score.

CP computerization and standardization

CPs are novel health care management plans that contain details of the interventions and procedures of patients’ treatment and follow-up. An example CP for ischemic stroke is shown in Figure 4. CPs are essential sources of data and data analytics in health care.^{26–28} However, a key factor that is impeding CP-based health data analytics and hindering the smooth transfer of CP data to other HISs is that CPs are prepared in hospitals without attention to standardizing their medical terms.^{26–28} After a thorough review of CP research found in literature, along with discussions with our domain experts at the Thunder Bay Regional Health Sciences Centre, it was clear that most CPs are currently developed using non standard local terms and

abbreviations.^{30–38} This situation makes CPs prone to human error and forms a challenge to exchanging them across medical institutions, thus limiting their adoption worldwide. This also causes the loss of valuable CP data because existing HISs use standardized terminology systems when encoding their medical terms. To give an example, the instruction “Complete Hx and PE” that is part of a CP for open appendectomy,³⁰ cannot be captured in HISs. To address this issue, we recently published a CP standardizing and digitizing framework²⁶ in which an instruction like this is standardized manually as “Complete History and Physical Examination” where “History and Physical Examination” is a standard SNOMED CT term whose SNOMED CT identification number is 53807006.²⁶ However, manual CP term standardization is a tedious and time-consuming process, and therefore, automation to CP term standardization was introduced in this research. CP computerization was the focus of many studies in the literature.^{8–16} For example, Liu et al.¹⁴ proposed an ontological approach for real-time CP monitoring. The main objective of their study was to establish communication between the CP and the electronic medical record system with the ability to display reminders on medical activities. The CP used in their prototype was related to unstable angina. Blaser et al.¹⁶ developed a prototype CP system that was embedded within an EMR system to guide patients’ treatment. For a more detailed and recent review, the reader is referred to Alahmar et al.²⁶ The literature review shows that the common theme in nearly all studies is that the goal of the computerization process was limited to connecting CPs with EMRs. We view this as a limitation because full CP digitization can only be achieved by standardizing their medical terms, which is an area of research that was not investigated in previous work related to CP computerization. This resulted in large amounts of CP data omitted in EMRs/HISs, and not utilized

Enter the SCTID: <input type="text" value="80146002"/>	
<input type="button" value="Search"/>	
<u>Contextualized questions:</u>	
Method Used:	<input type="checkbox"/> Drainage <input checked="" type="checkbox"/> Excision
Procedure site - indirect:	<input type="checkbox"/> Structure of abdominopelvic cavity <input type="checkbox"/> N/A
Procedure site - direct:	<input checked="" type="checkbox"/> Appendix structure
Direct substance:	<input type="checkbox"/> Substance <input type="checkbox"/> N/A
Priority:	<input type="checkbox"/> Emergency <input checked="" type="checkbox"/> Scheduled - priority <input type="checkbox"/> Routine <input checked="" type="checkbox"/> N/A
Access:	<input type="checkbox"/> Open approach - access <input type="checkbox"/> N/A
Using access device:	<input checked="" type="checkbox"/> Laparoscope, device <input type="checkbox"/> N/A
<input type="button" value="Find best matching SNOMED CT terms"/>	

Figure 9. Contextual relevant questions and expert's answers.

in data analytics. Our proposed solution for this challenge is to encode CP data using internationally recognized medical reference terminology (SNOMED CT). SNOMED CT encoding is realized by representing each CP medical term by its equivalent SNOMED CT term and SNOMED CT ID (SCTID). As mentioned above, this study introduces an automated CP standardization technique. Table 4 shows a comparison between this research and recent similar work in the literature.

CP standardization using semantic-based relatedness measure

To automate the CP standardization process, we propose a semantic-based search algorithm that exploits various

semantic relations. The algorithm supports physicians and domain experts in their task of standardizing CP terms. The main intuition is that the similarity between SNOMED CT terms can help the algorithm find and propose the accurate SNOMED CT terms while searching the hierarchy of terms within the SNOMED CT ontology. Thus, the use of semantic relations can help domain experts in the CP validation process by automating the process of finding the correct terms which are usually similar to the local terms. In a large number of cases, the inaccurate local term might be either a sibling or an abstract term of the desired term. For example, the correct and accurate term “Occlusive stroke”, whose SCTID is 373606000 is the subclass of the inaccurate term (i.e. initial term) “ischemic stroke”, whose SNOMED SCTID is 422504002.

Algorithm 1: CP SNOMED CT standardization using semantic score

Input : $SCTID_{init}$: Initial SCTID
Output: $SCTID_{approved}$: SCTID approved by domain experts

1 **Data Structures:**
2 CL : A hash-set to store the candidate SNOMED CT terms, initially empty
3 QOd : A dictionary to store the relevant question-options set based on the associated relations
4 $root$: A variable to store the $parent_of(SCTID_{init})$ based on is_a relation, or the expanded node, initially null
5 $SCTID_{approved} \leftarrow null$
/* Retrieve child terms & infer contextualized questions */

6 **repeat**
7 **foreach** $child\ c_i \in child_of(root)$ **do**
8 $QOd \leftarrow Get_Relations(c_i)$
9 Display relevant question and options
10 Collect expert's answers
11 **end**
12 $Threshold = \frac{\# Selected\ Options}{\# Selected\ Options + \# non\ Selected\ Relations}$
/* Display relevant SCT terms */
13 **foreach** $child\ c_i \in child_of(root)$ **do**
14 **if** ($similarity(SCTID_{init}, c_i) \geq Threshold$) **then**
15 $CL \leftarrow c_i$; // Append c_i to CL
16 **end**
17 **end**
/* Display CL to domain experts. Domain experts can either: expand a child node to check its children OR approve a child node as the standardized term */
18 **foreach** $child\ node\ c_i \in CL$ **do**
19 **if** (c_i is approved) **then**
20 $SCTID_{approved} \leftarrow c_i$
21 **else if** ($expand\ c_i$) **then**
22 $root \leftarrow c_i$
23 **end**
24 **end**

25 **until** ($SCTID_{approved} \neq null$) \vee ($CL = \emptyset$);

The algorithm is described below (see Algorithm 1). The algorithm is implemented in an interactive software system that assists domain experts in determining the accurate standardized SNOMED CT (SCT) terms. The algorithm limits the possible terms to only those that are semantically relevant to the initial term, and presents a set of contextualized relevancy questions. The contextualized relevancy questions are obtained from the non-taxonomic relations that each child SCT term has.

Figure 5 shows the SCT term “ischemic stroke (SCTID: 422504002)” with its child terms, and the non-taxonomic relations associated with it (i.e. “Finding site”, “Associated morphology”, and “Due to”). Based on each non-taxonomic relation of the child terms, a question set and options are presented to the domain experts to provide the appropriate context, as shown in Figure 6. Line 4 in Algorithm 1 initializes the root with the parent of the initial SCT term based on the semantic relation “is-a.” Then, iterative steps limit the candidate SCT terms to the most relevant and semantically similar terms to the initial term based on Eq. 3, where the F_1 -measure (i.e. the harmonic mean of the precision and recall, see equations (1) and (2)) represents the score of similarity between the expert’s selection of the contextualized options, and the semantic relations and associations of the child SCT terms. Note that the similarity score, presented in equation (3), for any given term could have a value of between [0–1], where zero means not contextually similar/related, and one means highly contextually related. For instance, based on the domain expert’s answers presented in Figure 6 and the non-taxonomic relations of “Occlusive stroke” and “Perinatal arterial ischemic stroke”, their similarity scores are determined as 0.91 and 0.6, respectively (Figure 7). Finally, the domain expert is presented with a limited number of plausible SCT terms to either select an appropriate term (as the approved standardized term) or expand a child term to repeat the process until the standardized term is found.

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

where TP are the true positive observations that represents the agreement between the expert’s selection and the existing semantic associations with the SCT term, and FP (false positive) are the semantic associations that are present in the child SCT term but are not selected by the expert (Figure 8).

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

where FN are the false-negative observations that represent the semantic associations that are selected by the expert but are not present in the child SCT term (Figure 8).

$$SimScore(SCT_1, SCT_2) = F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Figure 8 shows the use of F1 as a semantic score. The solid dots represent the options selected by the domain expert(s) and the hollow dots represent the contextualized options not selected by the domain expert(s).

Experimental results and discussion

Here, we present the experimental results and discussion for common CP standardization scenarios. The scenarios are related to local CP terms from Category 1 and Category 2 (as described in the introduction).

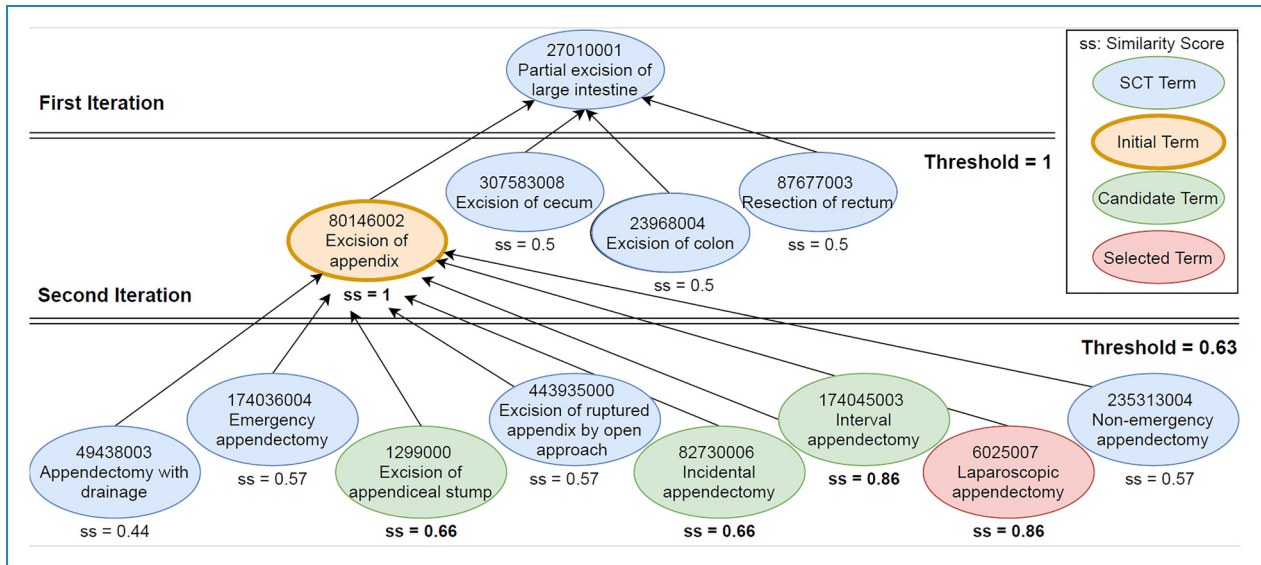


Figure 10. An illustration of selecting the appropriate SNOMED CT (SCT) term based on the similarity score.

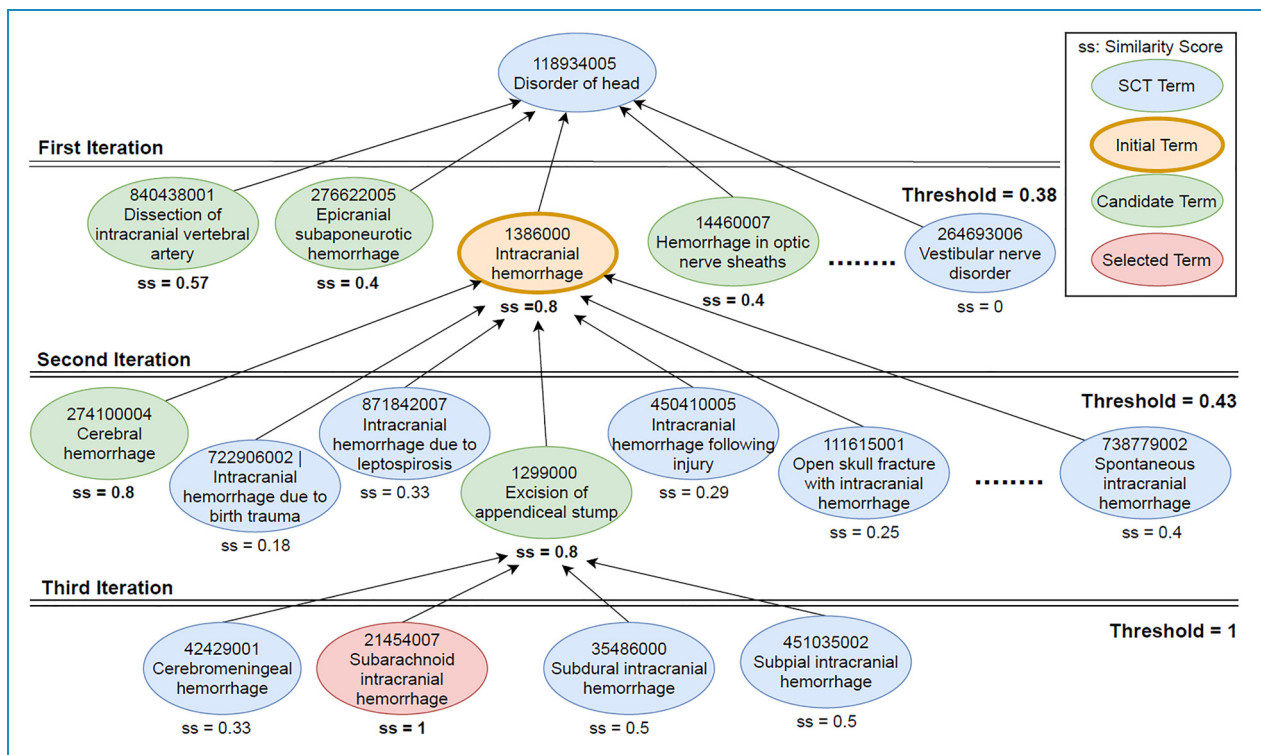


Figure 11. An illustration of selecting the appropriate SNOMED CT (SCT) term based on the similarity score.

Standardizing CP local terms from other terminology systems

In this scenario, we consider CP local terms that are used in terminology systems other than the SNOMED CT system (e.g. LOINC, ICD-10 and its variants like ICD-10-PCS, etc.). It should be noted that this is an example from

Category 2 local terms. The main challenge in converting these terms into standardized terms is finding the accurate SNOMED CT terms because terminology systems consist of many terms that look similar, but do not exactly convey the identical medical interventions or meanings. A typical scenario would be a term written in the CP from ICD-10 terminology, whereas the CP is required to be

Table 5. Search space reduction within each iteration.

Iteration#	Threshold	# SCTChildren	# SCT Candidates	% Search Space Reduction
1	0.375	104	4	96.2%
2	0.429	14	2	85.7%
3	1.0	4	1	75.0%

fully standardized using SNOMED CT prior to its inclusion in a hospital's IT system. An example is given below.

In this scenario, the ICD-10-PCS term “Resection of Appendix, Percutaneous Endoscopic Approach (0DTJ4ZZ)” was found in the CP. A straightforward similar term from SNOMED CT would be “Excision of appendix (SCTID: 80146002)” because both terms refer to similar medical procedures. For domain experts, the existence of “endoscopic approach” in the original term is important and must be reflected in the approved SCT term. The steps below outline how the accurate SNOMED CT term can be inferred based on the semantic relatedness approach. Note that the accurate term will be inferred to be “Laparoscopic appendectomy (SCTID: 6025007).” Initially, the algorithm starts by retrieving parent SCT term (parent of (80146002)) which is the SCT term “Partial excision of large intestine (SCTID: 27010001).” Then, for each child SCT term, the algorithm will retrieve the set of relevant questions with their respective answers based on all semantic relations associated with each child SCT term from the SNOMED CT ontology (Algorithm 1 lines 7–11). These lines form an iteration in the algorithm which is defined as a loop over the child SNOMED CT terms of the root SNOMED CT term under consideration. For the first iteration, the algorithm displays two contextually relevant questions based on the two existing relations (or attributes) possessed by child concepts (i.e. “Method” and “Procedure site - Direct”). All four child SCT terms, in this iteration, share the same option for the “Method” relation (i.e. “method: Excision - action”), however, they differ with the “Procedure site - Direct” relation (i.e. “Appendix structure”, “Cecum structure”, “Colon structure”, or “Rectum structure”). Based on the expert's response for the “Procedure site - Direct” question (“Appendix structure”), line 12 in Algorithm 1 (presented in Eq. 4) computes the threshold value. Then, for this iteration, out of the four child terms (“Excision of appendix”, “Excision of cecum”, “Excision of colon”, and “Resection of rectum”), the algorithm will display only the initial SCT term (i.e. “Excision of appendix”), as it is the most

Table 6. A dataset of initial and target SNOMED CT terms.

Initial Term	Target Term
(106063007) Cardiovascular finding (finding)	(1939005) Abnormal vascular flow (finding)
(230690007) Cerebrovascular accident (disorder)	(16371781000119100) Cerebellar stroke (disorder)
(230690007) Cerebrovascular accident (disorder)	(16661931000119102) Cerebrovascular accident due to stenosis of bilateral vertebral arteries (disorder)
(230690007) Cerebrovascular accident (disorder)	(329371000119101) Cerebrovascular accident due to occlusion of left middle cerebral artery by embolus (disorder)
(292671000119104) Cerebrovascular accident due to stenosis of left vertebral artery (disorder)	(16661931000119102) Cerebrovascular accident due to stenosis of bilateral vertebral arteries (disorder)
(301095005) Cardiac finding (finding)	(449543008) Paradoxical motion of ventricular septum (finding)
(371040005) Thrombotic stroke (disorder)	(444657001) Superior cerebellar artery syndrome (disorder)
(371041009) Embolic stroke (disorder)	(413758000) Cardioembolic stroke (disorder)
(422504002) Ischemic stroke (disorder)	(140921000119102) Ischemic stroke without coma (disorder)
(422504002) Ischemic stroke (disorder)	(373606000) Occlusive stroke (disorder)
(49601007) Disorder of cardiovascular system (disorder)	(241998008) Cardiovascular decompression injury (disorder)
(49601007) Disorder of cardiovascular system (disorder)	(95653008) Acute confusional migraine (disorder)
(62914000) Cerebrovascular disease (disorder)	(230716006) Carotid territory transient ischemic attack (disorder)
(62914000) Cerebrovascular disease (disorder)	(195200006) Carotid artery syndrome hemispheric (disorder)

Table 7. Search space reduction within each iteration.

Initial SCTID	Target SCTID	Iteration #	Threshold	# SCT Children	# SCT Candidates	% Search Space Reduction
106063007	1939005	1	0.125	92	1	98.9%
		2	0.143	45	1	97.8%
		3	0.286	45	36	20.0%
230690007	16371781000119100	1	0.273	94	33	64.9%
		2	0.400	51	42	17.6%
230690007	16661931000119100	1	0.273	94	33	64.9%
		2	0.375	51	39	23.5%
		3	1.000	1	1	0.0%
230690007	329371000119101	1	0.273	94	33	64.9%
		2	0.444	51	48	5.9%
		3	1.000	9	1	88.9%
		4	1.000	1	1	0.0%
292671000119104	16661931000119100	1	0.500	51	41	19.6%
		2	1.000	1	1	0.0%
301095005	449543008	1	0.286	45	3	93.3%
		2	0.375	83	11	86.7%
		3	1.000	1	1	0.0%
371040005	444657001	1	0.444	51	48	5.9%
		2	1.000	11	1	90.9%
		3	1.000	1	1	0.0%
371041009	413758000	1	0.600	51	39	23.5%
		2	1.000	1	1	0.0%
422504002	140921000119102	1	0.636	51	8	84.3%
		2	0.800	10	7	30.0%
422504002	373606000	1	0.600	51	7	86.3%
		2	0.833	10	4	60.0%
49601007	241998008	1	0.143	56	1	98.2%

(continued)

Table 7. Continued.

Initial SCTID	Target SCTID	Iteration #	Threshold	# SCT Children	# SCT Candidates	% Search Space Reduction
		2	0.300	48	34	29.2%
		3	0.667	5	1	80.0%
49601007	95653008	1	0.143	56	1	98.2%
		2	0.455	48	2	95.8%
		3	0.500	33	4	87.9%
128487001	230716006	1	0.364	48	3	93.8%
		2	0.333	33	4	87.9%
		3	0.750	8	1	87.5%
		4	0.800	3	2	33.3%
128487001	195200006	1	0.364	48	8	83.3%
		2	0.333	33	5	84.8%
		3	0.800	8	5	37.5%
		4	1.000	2	1	50.0%

relevant to the “Procedure site - Direct: Appendix structure,” where the expert will have the option to expand it and begin a second iteration. In the second iteration, by expanding the term (“Excision of appendix (SCTID: 80146002)”), a new set of contextually relevant questions are displayed to experts based on the new attribute relations of the child terms. Figure 9 shows these contextually relevant questions and the selected answers chosen by the domain experts. Based on these questions and answers, a new similarity score threshold value is computed using equation (4). Finally, the SCT terms with a similarity score above or equal to the current iteration’s threshold value are added to the candidate list of terms to be displayed to the expert and ranked according to their similarity score, from highest to lowest. Finally, the expert decides on the approved standardized SCT term, which is “Laparoscopic appendectomy (ID: 6025007),” with a similarity score of 0.86. Figure 10 shows the iteration process resulting from Algorithm 1, showing the similarity scores for each SCT term and the threshold value for each iteration. The threshold value is defined in Eq. 4 as the number of options selected by the expert divided by the sum of selected options and nonselected relations. This ensures that the threshold calculated value is a reflection of the expert’s minimum agreement with the presented options.

$$Threshold = \frac{SO}{SO + NSR}, \quad (4)$$

where SO is the number of selected options and NSR is the number of nonselected relations.

Standardizing CP local terms from SNOMED CT

The next scenario relates to an initial SNOMED CT term “Intracranial hemorrhage (SCTID: 1386000)” and an approved term “Subarachnoid intracranial hemorrhage (SCTID: 21454007).” Both terms are used to encode hemorrhagic stroke, a common medical condition. The subtle difference between the terms is that intracranial hemorrhage refers to bleeding that occurs when a blood vessel within the skull is ruptured or leaks, whereas subarachnoid intracranial hemorrhage is a type of stroke caused by bleeding into the space surrounding the brain (not within the skull). Similar to the first scenario, Figure 11 shows the iterative process described in Algorithm 1 for selecting the standardized approved SNOMED CT term, which is, in our case, “subarachnoid intracranial hemorrhage (SCTID: 21454007)” given the initial term “intracranial hemorrhage (SCTID: 1386000).”

Table 5 shows the details of each of the three iterations required in this case, including the percentage reduction of SNOMED CT search space after each iteration. In iteration 1, among 104 SNOMED CT child terms, the algorithm utilizes the expert’s answers for the contextualized questions to reduce the candidate correct terms to only 4 possible terms, achieving a 96.2% reduction in the search space.

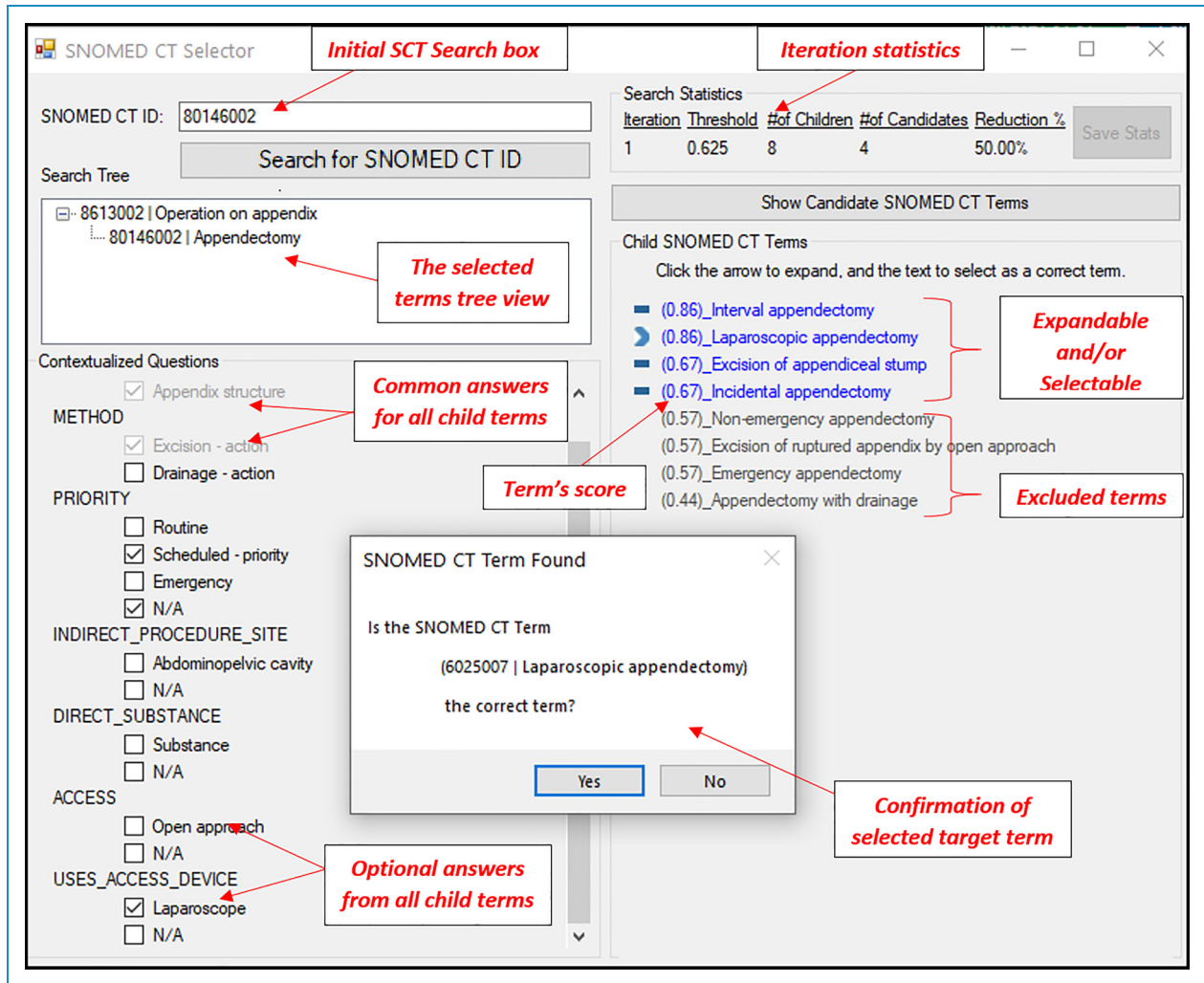


Figure 12. The main screen of the developed software application (SNOMED CT selector). SNOMED CT: Systematized Nomenclature of Medicine – Clinical Terms.

Iteration 2 finds 14 child terms and limits the candidate terms to 2, achieving 85.7% search space reduction. The correct SNOMED CT term “subarachnoid intracranial hemorrhage (SCTID: 21454007)” was located in iteration 3 among the four existing child terms. This scenario demonstrates the capability of the proposed algorithm in significantly reducing the search space for the semantically relevant SNOMED CT terms.

To demonstrate the robustness of the proposed algorithm, a dataset of 14 pairs of SNOMED CT terms was examined, where the domain experts determined a target SNOMED CT term, given an initial local CP term, as shown in Table 6. Based on the responses of the experts to the contextualized questions for each pair, the algorithm demonstrates a consistent search space reduction and convergence to the target term for all scenarios, as shown by the experimental results in Table 7. The experimental results show that the search space reduction reached 98.9% on individual iterations and an overall

average reduction of 71.3% considering the largest search tree in each iteration (i.e. the search tree that has the largest number of children). The proposed algorithm was implemented in a prototype software system that we made available on GitHub*. The developed software (with its graphical user interface (GUI)) enables domain experts to search for a SNOMED CT term by ID or name. Then, based on its children’s semantic relations within SNOMED CT ontology, the system presents all possible relations and options, as shown in Figure 12. Once the expert selects the relevant answers, the system filters out child SNOMED CT terms with scores below the threshold calculated using equation (4). This, in turn, allows for further expansion and searching within subsequent children, or selecting a child term as a standardized target term that is approved by domain experts. Figure 12 shows the main screen of the developed software system and outlines its components as described above.

It is worth mentioning that this research is not proposing a new standard but rather adopting an existing international standard. SNOMED CT is a well-known standard that has been successfully used and maintained worldwide since 1999. This contributes to the success and adoption of the proposed method in this research.

Conclusion and future work

Automation in CP standardization must be addressed carefully due to patient safety considerations associated with the medical data stored in CPs. In this research, we demonstrated a new automated CP standardization approach that achieves a balance between machine intervention and domain experts' control. Experimental results reveal that the proposed method achieves a high percentage of SNOMED CT search space reduction, which saves domain experts time and helps them to efficiently locate the correct standardized CP terms. The developed standardization approach can be used in other ontologically represented domains like standardizing clinical/medical terms in EMRs and medication prescriptions. The emphasis in this research was on the standardization of initial local CP terms that exist in SNOMED CT, but are not accurate based on domain expert's judgment. Therefore, a key task in the presented approach was to apply semantic relatedness to search the SNOMED CT ontology for terms with semantic relatedness to the initial SNOMED CT terms. The main limitation of this research is that local CP terms or abbreviations that are nonexistent in any terminology system were not addressed. Thus, our future investigation in this domain will concentrate on standardizing this type of local CP terms. In such cases, the initial term cannot be found within SNOMED CT ontology. Thus, standardizing the terms requires matching the local term to an external knowledge base. Such an approach would involve the use of NLP techniques different than the technique used in this study (such as word embedding methods).

Acknowledgments: We extend a sincere thank you to the Thunder Bay Regional Health Sciences Centre (TBRHSC) and The Ottawa Hospital for providing us with sample CPs. We also wish to thank our domain experts, Dr A. Hassan and Dr K. Darko, as well as the nurses from the Regional Stroke Unit, who helped us throughout our CP automation research.

Contributorship: All authors contributed equally to the research and approved the final version of the manuscript.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: Not required.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Guarantor: AA.

Informed consent: Not applicable, because this article does not contain any studies with human or animal subjects.

ORCID iD: Ayman Alahmar  <https://orcid.org/0000-0003-4011-1023>

Trial registration: Not applicable, because this article does not contain any clinical trials.

*https://github.com/mohannad57/SNOMED_CT_Algo.

References

1. Al-Ashwal RH and Supriyanto E. Clinical pathway in cardiovascular disease management. In Dewi D, Hau Y, Khudzari A, Muhamad I, Supriyanto E, editors. *Cardiovascular Engineering. Series in BioEngineering*. Singapore: Springer, 2020. doi:10.1007/978-981-10-8405-8_6
2. Zhang C and Xiao J. Application of fast-track surgery combined with a clinical nursing pathway in the rehabilitation of patients undergoing total hip arthroplasty. *J Int Med Re* 2020; 48: 0300060519889718.
3. Jabbour M, Newton AS and Johnson D. Defining barriers and enablers for clinical pathway implementation in complex clinical settings. *Implement Sci* 2018; 13: –3.
4. Alahmar A, Mohammed E and Benlamri R. Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes. In 2018 *4th International Conference on Big Data Innovations and Applications (Innovate-Data IEEE)* 6 August 2018 (pp. 38–43). doi:10.1109/Innovate-Data.2018.00013
5. Rathore MM, Ahmad A and Paul A. Real-time medical emergency response system: exploiting IoT and big data for public health. *J Med Syst* 2016; 40: 1–10.
6. Gul MJ, Subramanian B and Paul A. Blockchain for public health care in smart society. *Microprocess Microsyst* 2021; 80: 103524.
7. Sangal RB, Liu RB and Cole KO. Implementation of an electronic health record integrated clinical pathway improves adherence to COVID-19 hospital care guidelines. *Am J Med Qual* Jan. 11, 2022. doi:10.1097/jmq.0000000000000036
8. Katzan IL, Fan Y and Speck M. Electronic stroke CarePath: integrated approach to stroke care. *Circ: Cardiovasc Qual Outcomes* 2015; 8: S179–S189.
9. Abidi SR and Abidi SS. An ontological modeling approach to align institution-specific Clinical Pathways: Towards inter-institution care standardization. In 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS) 20 June 2012 (pp. 1–4).

10. Abidi SR, Abidi SS, Butler L, et al. Operationalizing prostate cancer clinical pathways: An ontological model to computerize, merge and execute institution-specific clinical pathways. In Workshop on Knowledge Management for Health Care Procedures 21 July 2008 (pp. 1–12). Springer, Berlin, Heidelberg.
11. Daniyal A, Abidi SR and Abidi SS. Computerizing clinical pathways: ontology-based modeling and execution. In Medical Informatics in a United and Healthy Europe 2009 (pp. 643–647). IOS Press.
12. Fudholi DH and Mutawalli L. An ontology model for clinical pathway audit. In 2018 4th International Conference on Science and Technology (ICST) 7 August 2018 (pp. 1–6).
13. Gibbs J, Sutcliffe LJ and Gkatzidou V. The eClinical care pathway framework: a novel structure for creation of online complex clinical care pathways and its application in the management of sexually transmitted infections. *BMC Med Inform Decis Mak* 2016; 16: 1–9.
14. Liu J, Huang Z and Lu X. An ontology-based real-time monitoring approach to clinical pathway. In 2014 7th International Conference on Biomedical Engineering and Informatics 14 October 2014 IEEE (pp. 756–761).
15. Ye Y, Jiang Z and Diao X. An ontology-based hierarchical semantic modeling approach to clinical pathway workflows. *Comput Biol Med* 2009; 39: 722–732.
16. Blaser R, Schnabel M and Biber C. Improving pathway compliance and clinician performance by using information technology. *Int J Med Inf* 2007; 76: 151–156.
17. SNOMED International. <https://confluence.ihtsdotools.org/>, Accessed: 2 February 2022.
18. Clinical pathways and order sets in Epic, The Ottawa Hospital, <https://www.ottawahospital.on.ca/en/>, Accessed: 1 January 2022.
19. Logical Observation Identifiers, Names, and Codes (LOINC), <https://loinc.org/>, Accessed: 2 February 2022.
20. International Classification of Diseases, Tenth Revision (ICD-10), <https://www.cdc.gov/nchs/icd/icd10.htm>, Accessed: 2 February 2022.
21. The ICD-10 Procedure Coding System (ICD-10-PCS), <https://www.cms.gov/Medicare/Coding/ICD10/2020-ICD-10-PCS>, Accessed: 2 February 2022.
22. Harispe S, Ranwez S and Janaqi S. Semantic similarity from natural language and ontology analysis. *Synth Lect Hum Lang Technol* 2015; 8: 1–254.
23. Feng Y, Bagheri E and Ensan F. The state of the art in semantic relatedness: a framework for comparison. *Knowl Eng Rev* 2017; 32: 1–30. doi:10.1017/S0269888917000029
24. AlMousa M, Benlamri R and Khoury R. Exploiting non-taxonomic relations for measuring semantic similarity and relatedness in WordNet. *Knowl Based Syst* 2021; 212: 106565/ 1–19. doi: 10.1016/j.knosys.2020.106565
25. Liu H, Bao H and Xu D. Concept vector for semantic similarity and relatedness based on wordnet structure. *J Syst Softw* 2012; 85: 370–381.
26. Alahmar A, Crupi ME and Benlamri R. Ontological framework for standardizing and digitizing clinical pathways in healthcare information systems. *Comput Methods Programs Biomed* 2020; 196: 105559. doi:10.1016/j.cmpb.2020.105559
27. Alahmar A and Benlamri R. SNOMED CT-based standardized e-clinical pathways for enabling big data analytics in healthcare. *IEEE Access* 2020; 8: 92765–92775. doi:10.1109/ACCESS.2020.2994286
28. Alahmar A and Benlamri R. Optimizing Hospital Resources using Big Data Analytics with Standardized e-Clinical Pathways. In 2020 IEEE Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCOM/CyberSciTech) 17 August 2020 (pp. 650–657). doi:10.1109/DASC-PiCom-CBDCOM-CyberSciTech49142.2020.00112
29. Canada Health Infoway, <https://infoway-inforoute.ca/en/>, Accessed: 2 February 2022.
30. Hilario AL, Oruga JD and Turqueza MP. Utilization of clinical pathway on open appendectomy: a quality improvement initiative in a private hospital in the Philippines. *Int J Health Sci* 2018; 12: 43.
31. Ibeziako P, Brahmabhatt K and Chapman A. Developing a clinical pathway for somatic symptom and related disorders in pediatric hospital settings. *Hosp Pediatr* 2019; 9: 147–155.
32. Patel NA, Bly RA and Adams S. A clinical pathway for the postoperative management of hypocalcemia after pediatric thyroidectomy reduces blood draws. *Int J Pediatr Otorhinolaryngol* 2018; 105: 132–137.
33. Ovaere S, Boscart I and Parmentier I. The effectiveness of a clinical pathway in liver surgery: a case-control study. *J Gastrointest Surg* 2018; 22: 684–694.
34. Lanig IS, New PW and Burns AS. Optimizing the Management of Spasticity in people with spinal cord damage: a clinical care pathway for assessment and treatment decision making from the ability network, an international initiative. *Arch Phys Med Rehabil* 2018; 99: 1681–1687.
35. Shubert CR, Kendrick ML and Habermann EB. Implementation of prospective, surgeon-driven, risk-based pathway for pancreatoduodenectomy results in improved clinical outcomes and first year cost savings of \$1 million. *Surgery* 2018; 163: 495–502.
36. Wylde V, Bertram W and Beswick AD. Clinical-and cost-effectiveness of the STAR care pathway compared to usual care for patients with chronic pain after total knee replacement: study protocol for a UK randomised controlled trial. *Trials* 2018; 19: 1–3.
37. Li M, Zhang J and Gan TJ. Enhanced recovery after surgery pathway for patients undergoing cardiac surgery: a randomized clinical trial. *Eur J Cardiothorac Surg* 2018; 54: 491–497.
38. Kamil W, Hian OY and Mohd-Said S. Development of clinical pathway for non-surgical management of chronic periodontitis. *Malaysian Journal of Public Health Medicine* 2018; 2018: 26–32.
39. Norra C, Ueberberg B and Juckel G. A new electronically based clinical pathway for schizophrenia inpatients: a longitudinal pilot study. *Schizophrenia Research, Elsevier* 2021; 238: 82–90.