Journal of Biomedical Informatics, Volume 111, November 2020, 103585.
This is the accepted manuscript version of the article made available under the CC BY-NC-ND 4.0 license,
the final published journal article is available at: https://doi.org/10.1016/j.jbi.2020.103585.

# Automatic Full Conversion of Clinical Terms into SNOMED CT Concepts

**Rohit J. Kate**[1]
Department of Computer Science
University of Wisconsin-Milwaukee
Milwaukee, WI, USA

## Abstract

SNOMED CT is the most comprehensive clinical ontology and is also amenable for automated reasoning. However, in order to unleash its full potential for automated reasoning over clinical text, a mechanism to convert clinical terms into SNOMED CT concepts is necessary. In this paper we present, to the best of our knowledge, the first such complete conversion method that is also capable of converting clinical terms into post-coordinated concepts which are not already listed in SNOMED CT. The method does not require any additional manual annotations and learns only from existing SNOMED CT terms paired with their concepts. The method is based on identifying the defining relations of the clinical concept expressed by a clinical term. We evaluate our method on a large-scale using existing data from SNOMED CT as well as on a small-scale using manually annotated dataset of clinical terms found in clinical text.

**Keywords**
clinical terms; SNOMED CT; machine learning; ontology

## 1. Introduction

SNOMED CT [1] is the most comprehensive clinical ontology in the world which includes more than 300,000 concepts that have more than 1.5 million relations between them. It is based on the description logic framework [2] which makes it amenable for automated reasoning. Because concepts are defined in terms of their relations with other concepts in SNOMED CT, by simply identifying a concept in this ontology one can know a lot about it through its explicitly stated relations as well as through its implicitly inferred relations. However, given that a large portion of electronic health record data is in the form of text [3], in order to unleash its full potential for automated applications, a method is needed to automatically convert clinical terms into concepts in SNOMED CT. For example, if a clinical text mentions the clinical term “meningitis caused by virus” then by converting it into its respective SNOMED CT concept, one would know through its explicitly stated relations that it is an infectious disease caused by virus whose finding site is meninges structure and that it leads to inflammation. Furthermore, one can also implicitly infer through hierarchical relations of meninges structure that it is an infection of the nervous system. This information can then help with automated applications, for example, a search for retrieving records of all patients whose nervous systems have been infected. For this task, a simple keyword search, such as “nervous system infected”, will not otherwise retrieve the patient record that has “meningitis caused by virus” in its clinical text because the keyword search will not match that term.

However, finding the SNOMED CT concept corresponding to a clinical term mentioned in text is not trivial. This is because although each concept in SNOMED CT has been listed with its few names,

---

[1] Corresponding author.
*Email-address:* katerj@uwm.edu

this often does not include all its possible names due to the variability present in natural languages. For example, “meningitis caused by virus” is, in fact, not listed in SNOMED CT, although its synonyms “abacterial meningitis” and “viral meningitis” are listed. Hence simple string matching does not always work to find the concept in SNOMED CT corresponding to a clinical term.

Besides a clinical term not being listed in SNOMED CT, sometimes the concept itself may not be present in SNOMED CT. For example, not just the term but even the concept corresponding to the term “acute abdominal pain on the right side” is not present in SNOMED CT. Clinical text annotators have noted that 27.8% of clinical terms of disorder semantic type [4] and 19.75% of clinical terms of all semantic types [5] mentioned in clinical text did not have their corresponding concepts present in SNOMED CT. It is, in fact, nearly impossible to include all possible clinical concepts in a clinical ontology because clinical concepts can always be composed using other clinical concepts which can lead to an explosion in the number of clinical concepts; for example, consider the following list of clinical terms – “pain in the arm”, “pain in the left arm”, “pain in the upper left arm”, “acute pain in the upper left arm”, “acute pain in the upper left arm due to ischemia”, “history of the acute pain in upper left arm due to ischemia”, etc. A list of such compositionally created concepts can thus go on endlessly and exponentially more concepts can be created by varying the component concepts. Hence no clinical ontology is expected to include all possible clinical concepts even though many of the concepts that are not included in it may be found in clinical text. For this reason, SNOMED CT provides a mechanism through which one can create new concepts by identifying its relations with existing SNOMED CT concepts.

For all the above reasons, it is necessary to have a method which can automatically convert clinical terms into concepts in SNOMED CT, including to new concepts. Figure 1 shows an example of the conversion task for the clinical term “acute abdominal pain on the right side”. Although not an uncommon clinical term, it is not present in the existing SNOMED CT. However, it can be converted into a SNOMED CT concept by identifying its defining relations with existing SNOMED CT concepts, as depicted in Figure 1. In this work, we present, to the best of our knowledge, the first such complete conversion method. Given that a concept in SNOMED CT is represented in terms of its relations with other concepts, our method semantically converts a clinical term into its ontological concept by identifying the related concepts that define it. The method does not require any new manual annotations and learns only from pairs of clinical terms and their corresponding concepts already present in SNOMED CT. We evaluated the method using existing data from SNOMED CT to measure its performance on converting clinical terms to their concepts. We also did a small-scale manual evaluation using some clinical terms found in real clinical text that did not have concepts present in the existing SNOMED CT.

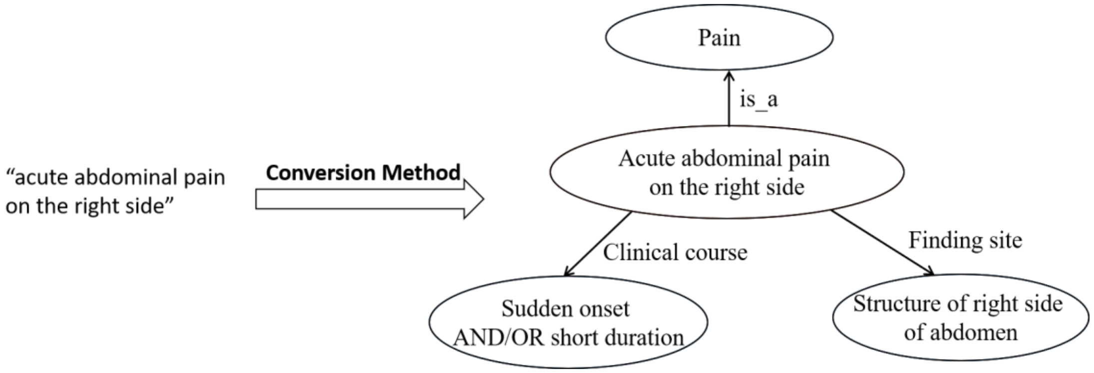

Figure 1. An example of converting a clinical term into its SNOMED CT concept by identifying its relations with existing SNOMED CT concepts. The concepts are shown as nodes in the graph and the relations are shown as edges connecting them.

The rest of the paper is organized as follows. We present past related work in the next section. This is followed by a section that provides some background on SNOMED CT which is necessary to describe our method. We describe our method in Section 4 followed by the results and discussion in Section 5. The last section mentions a few limitations of our work and avenues for future work. In the rest of the paper, whenever we want to refer to the clinical concept behind a clinical term then we will write the clinical term in italics and without double quotes, and whenever we want to refer to the clinical term itself then we will write it in double quotes. For example, when we want to refer to its clinical concept we will write *meningitis caused by virus* but when we want to refer to the clinical term itself (i.e. the natural language string) then we will write “meningitis caused by virus”.

## 2. Related Work

Most early related work had focused on automatic coding of clinical terms found in clinical text to their respective concepts in SNOMED CT [6,7,8,9,10]. These approaches either exactly or approximately matched clinical terms to their names in the ontology or used approaches similar to information retrieval. Recently, there has been interest in the clinical term normalization task in which clinical terms extracted from clinical text need to be assigned their respective clinical concept identifiers from an ontology such as SNOMED CT. Clinical term normalization was part of the larger shared-tasks of ShARe/CLEF 2013 Task 1 [11] and SemEval 2014 Task 7 [4] while the 2019 n2c2 Challenge Track 3 [12] was exclusively for clinical term normalization. Besides exact or approximate matching methods, researchers have also developed various machine learning based methods for clinical term normalization [13,14,15].

We point out that although related, coding or clinical term normalization is very different from the conversion task that is the focus of this paper. Clinical term normalization is restricted to terms whose concepts are already present in the ontology and typically requires matching between the clinical term and one of the concept’s names in the ontology. If a clinical term’s concept is not already in the ontology then the normalization systems are expected to simply output so (in the shared-tasks the systems were to output “CUI-less”). In contrast, the task being considered in this paper requires semantically mapping clinical terms to their concepts by identifying their relations and related concepts and in doing so generate new concepts for clinical terms if they are not already present in SNOMED CT. In the manually annotated MCN corpus [5] used in the 2019 n2c2 Challenge Track 3 [12], if a clinical term’s concept was not already present in SNOMED CT, its subterms were normalized instead of the entire term. For example, there is no concept in SNOMED CT corresponding to the term “left breast biopsy”, hence in the MCN corpus the subterms “left” and “breast biopsy” were annotated with their SNOMED CT concept identifiers. In contrast, in the task being considered in this paper, our method is expected to generate a new SNOMED CT concept for “left breast biopsy”.

There has been also work in automatically generating formal ontological definitions from textual definitions, such as “Baritosis is pneumoconiosis caused by barium dust” [16]. This task is, however, different from our task because in our task the input is a clinical term (i.e. a phrase) that expresses the concept (Figure 1 shows an example), whereas in their task the input is an entire sentence that defines the concept. A pattern-based method was presented in [17] to construct a possible SNOMED CT concept from concepts extracted from text, for example, if the text is “Given the rapid extension of the subgaleal bleeding, coagulopathy workup was initiated”, then the extracted concepts will be {*subgaleal area*, *bleeding* and *blood coagulation panel*}.This is also different from our task in which the input is a clinical term.

The closest work that had considered the clinical term to SNOMED CT conversion task is our own previous work [18]. In that work, as an important step towards the goal of full conversion, we had introduced the task of relation identification. The goal of the relation identification task is to classify whether two given clinical terms are related by a particular SNOMED CT relation or not. For example, given two clinical terms, say, “gastric ulcer” and “stomach structure”, whether they are related by the *finding-site* relation or not. This is a binary classification task and for each relation we had built a separate machine learning based binary classifier. These classifiers were trained using support-vector machines (SVM) [19] with a kernel that we had specifically designed for this task. The training for the classifier for each relation was done using total 2500 randomly selected positive examples of clinical term pairs of concepts that are known to be related by the relation, and 2500 randomly selected negative examples of clinical term pairs of concepts that are not related by the relation. The testing was done on similarly selected same number of positive and negative examples, and the whole process was repeated five times. The results showed good performance on the relation identification task. However, these classifiers were then not applied to the full conversion task in which for a given clinical term all its related concepts need to be determined. Unlike the relation identification task in which the second clinical term is already given and the classifier only needs to determine if it is related or not, in the full conversion task, no second clinical term is already given, instead, the method needs to determine the related concept on its own out of the thousands of concepts in SNOMED CT. In a pilot study within a small subset of SNOMED CT, it was found that applying a relation identification classifier one-by-one to all possible clinical terms would identify the correctly related concepts with only 10-20% maximum F-measure [18]. Furthermore, the SVM-based learning method was also computationally intensive and could not scale beyond a few thousand training and test examples. In this paper, we present a complete method for the full conversion of clinical terms to SNOMED CT concepts and also evaluate its performance on this full conversion task. We also present a new and faster method for identifying related concepts that easily scales up to the size of the entire SNOMED CT.

Recently, deep learning based classifiers were introduced [20] for the relation identification task and the results were compared with that of [18]. The authors also applied their classifiers to identify the relations present between the given clinical term and its subterms as identified by the MetaMap system [21]. But like [18], they did not present results on the full conversion task. It should be also pointed out that not all related concepts are present as subterms within a term; for example, “stomach structure” is not present as a subterm in “gastric ulcer” even though it is related. Hence their method may not be able to find all the related concepts. The full conversion task requires finding connections between the words present in clinical terms and the related concepts; for example, it needs to learn that the presence of word “gastric” is an indicative of the *stomach structure* concept. Besides being a method for the full conversion task, another reason the method presented in this paper is different from the methods presented in [18] and [20] is that it relates words in the clinical term directly to the related concepts (i.e. not just indirectly to the clinical term of the related concept). For example, it would directly relate word “gastric” to concept *stomach structure*, not just indirectly to the clinical term “stomach structure”. This is important because the words of “stomach structure” could also be found in other clinical terms, such as, “stomach wall structure” or “entire mouth, esophagus and stomach structures”, which can potentially lead to errors in identifying the related concept. We show through an ablation study later in the paper that the component of our method that directly related to the concepts was more important than the component that indirectly related through the clinical terms. We also point out that no related work had utilized the distinction between primitive and fully-defined concepts or leveraged normal forms of concepts which we describe in the next section.

## 3. SNOMED CT Background

This section provides a brief background on SNOMED CT which is necessary to describe our method. For more details on SNOMED CT one may refer to its reference sources such as [22] or [23].

### 3.1. Concepts and Relations

Every concept in SNOMED CT stands for a unique medical concept and has a unique identifier and a few synonymous names which are also called descriptions. One of these names is called full-specified name which is intended to be as unambiguous as possible, stable across multiple contexts, and optimally understandable [24], although it may not be the clinically preferred term for the concept. SNOMED CT includes descriptions in multiple languages but for this work we only used descriptions in English. Each concept also has a semantic type out of SNOMED CT's nineteen topmost level hierarchies that include, for example, *disorder*, *finding*, *body structure*, etc. The semantic type of a concept is indicated inside parenthesis with its fully-specified name. Figure 2 shows an example of the clinical concept of *viral meningitis* whose unique identifier is 58170007, fully-specified name is "viral meningitis", semantic type is *disorder*, and two other descriptions are "abacterial meningitis" and "aseptic meningitis, viral".

A concept in SNOMED CT is represented in terms of its relations with other concepts. An important relation is *is-a* which relates the concept to its parent, or in other words indicates – the given concept is a subtype of which other concept in SNOMED CT. There can be multiple *is-a* relations for a concept. The remaining relations are attribute relations which are also always true for the concept, for example, *finding-site*, *causative agent*, etc. Each relation in SNOMED CT has a domain semantic type and a range semantic type based on the semantic types of the concepts it relates. For example, the relation *causative agent* has the domain semantic type of *disorder* and the range semantic type of *organism*. A relation by the same name can have different domains and ranges; for example, the relation *finding-site* can be between *disorder* and *body structure* as well as between *finding* and *body structure*. In this work, we treat such cases as separate relations and indicate each relation along with its domain and range in parenthesis to distinguish them if needed, for example, *finding-site*(*disorder*, *body structure*) and *finding-site*(*finding*, *body structure*). It should be noted that *is-a* relations have the same semantic types for both the domain and the range, for example, *is-a*(*disorder*, *disorder*) and *is-a*(*finding*, *finding*). Other clinical aspects about a concept, such as symptoms or treatments, are not included in SNOMED CT because they may not be always true for the concept (for example, a symptom of *viral meningitis* may not be universally present in all patients of *viral meningitis*). Figure 2 shows the relations for the concept *viral meningitis* in the form of a graph where the names of the relations are shown on the edges and the related concepts are shown as nodes. For brevity, only the full-specified names and semantic types are shown for the related concepts. We have shown the concept and its relations in the form of a graph for visual ease. SNOMED CT otherwise also comes with a compositional grammar in which a concept can be written using the syntax of the grammar.

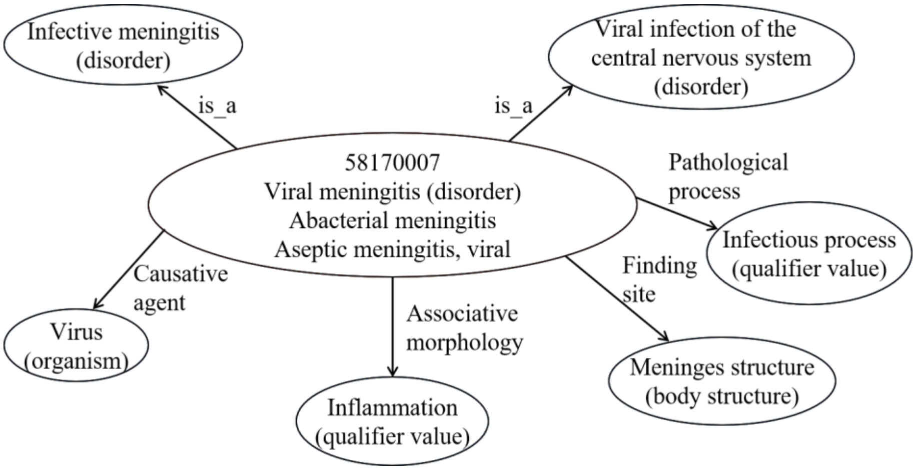

Figure 2. An example SNOMED CT concept *viral meningitis* shown with its related concepts.

### 3.2. Primitive and Fully-Defined Concepts

In SNOMED CT, concepts are represented in terms of their relations with other concepts. But a distinction is made between two types of concepts. The concepts that can be uniquely identified through its relations are called *fully-defined,* otherwise they are called *primitive*. Figure 3 (a) shows the two relations of the concept *asthma* in SNOMED CT. *Asthma* has *is-a* relation with *disorder of respiratory system* and *finding-site* relation with *airway structure*. However, these two relations cannot uniquely identify *asthma*; for instance, *upper airway resistance syndrome* (Figure 3 (b)) also has the exact same two relations with the exact same two concepts. Thus there is another disorder of respiratory systems that has airway structure as the finding site. Hence *asthma* is a primitive concept, and so is *upper airway resistance syndrome*. In other words, these concepts are incompletely defined in SNOMED CT and hence are called primitive.

Figure 3 (c), on the other hand, shows the relations of the concept *substance induced asthma*. It has *is-a* relation with *asthma* and in addition, it has *causative-agent* relation with *substance*. These two relations uniquely identify *substance induced asthma* because any *asthma* caused by some *substance* can only be *substance induced asthma*. Hence *substance induced asthma* is a full-defined concept. All concepts present in SNOMED CT are marked to be either primitive or fully-defined. About 75% of concepts in SNOMED CT are primitive while the rest are fully-defined.

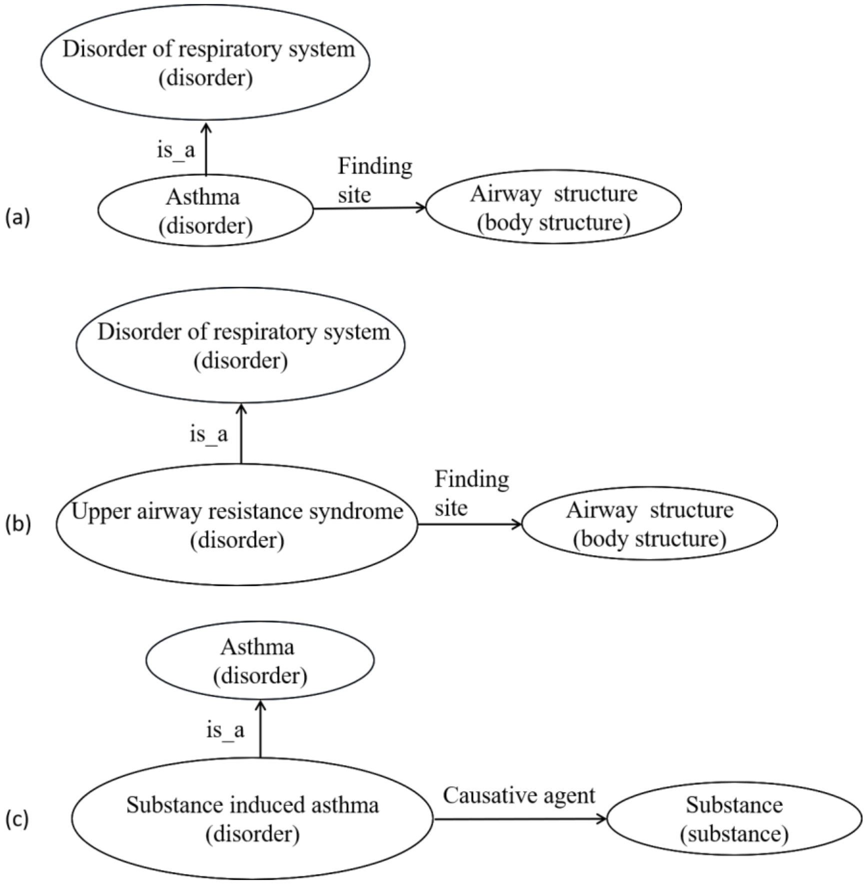

Figure 3. (a) *Asthma* and (b) *upper airway resistance syndrome* are primitive concepts that cannot be uniquely identified by their relations alone (the two, in fact, have identical relations). In contrast, (c) *substance induced asthma* is a fully-defined concept that can be uniquely identified from its relations.

### 3.3. Normal Form of Concepts and Concept Normalization

In SNOMED CT, there is often more than one way to represent a concept in terms of its relations. For example, Figure 4 shows two different ways in which *severe pain in the stomach* can be represented using its relations with other concepts. In Figure 4 (a), it is represented as a subtype of *abdominal pain* which is *severe*. In contrast, in Figure 4 (b), it is represented as a subtype of *pain* which is *severe* and whose *finding-site* is *stomach structure*. Please note that both the representations are, in fact, equivalent because *abdominal pain* itself is a subtype of *pain* whose *finding-site* is *stomach structure*. However, in order to automatically test equivalence or subsumption between two concepts, it is imperative to have a standard form for representing concepts that is unique for every concept. Such a standard way of representing a concept in SNOMED CT is called its *normal form*.

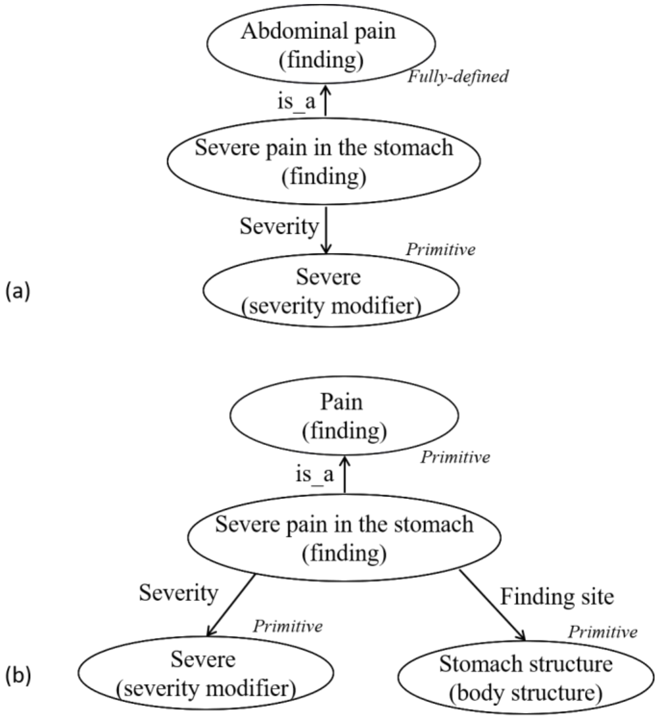

Figure 4. (a) The concept *severe pain in the stomach* is not in its normal form because all the concepts representing it are not primitive. (b) The same concept is in its normal form because all the concepts representing it are primitive.

A representation form for a concept is called normal if all the concepts present in it are primitive concepts. This definition makes the normal form unique for every concept, that is, there may be multiple ways to represent a concept but its normal form will be exactly one. The form shown in Figure 4 (a) is not a normal form because the related concept *abdominal pain* is not a primitive concept but is a fully-defined concept as a *pain* whose *finding-site* is *stomach structure*. But the form shown in Figure 4 (b) is a normal form because all the related concepts *pain*, *stomach structure* and *severe* are primitive. The forms of concepts that come with SNOMED CT (i.e. upon downloading its files) are not typically in normal form and are said to be in their *distributed form*. Distributed form is preferred over normal form in SNOMED CT's distribution files because sometimes normal forms are harder for humans to grasp. However, normal form is preferable for automated processing that involves checking for equivalence or subsumption between concepts.

There are two types of normal forms in SNOMED CT *long normal form* and *short normal form* which are, in fact, inter-convertible. In the long normal form of a concept, all the non-subtype relations of its parent concepts are included but in the short normal form they are not included. The form shown in Figure 3 (c) for *substance induced asthma* is its short normal form. Its long normal form will also include *asthma's* non-subtype relation *finding-site* to *airway structure*. The distinction between the two normal forms is made for computational efficiency in subsumption checking. But given that in this work we are not doing subsumption checking, we only use the short normal form which is sufficient for our purpose. As later described in Section 4, we use short normal forms of concepts for training as well as for evaluating our method. In the rest of the paper, by normal form of a concept we will always mean its short normal form. We point out that the short normal form of a primitive concept is considered to be the concept itself. An algorithm to automatically convert a SNOMED CT concept into its normal form is given in SNOMED CT's Terminology Services Guide [22] .

We want to point out that concept normalization described above is very different from term normalization. In term normalization, a term is given and one needs to determine its corresponding concept in a terminology or an ontology; for example, “left atrium dilated” normalizes to “left atrium dilation” in SNOMED CT. Term normalization is typically done by either exactly or approximately matching the term with one of the descriptions of the concept. It does not require analyzing the relations of the concept. The n2c2 Challenge Track 3 [12] and SemEval 2014 Task 7B [4] were both term normalization tasks. The recently developed MCN corpus [5] is, in fact, a corpus for term normalization, although it has been named “medical concept normalization” (MCN). On the other hand, concept normalization does not involve any clinical term at all and only involves rewriting the relations that define the concept.

### 3.4. Post-Coordinated Concepts

A predefined concept already present in SNOMED CT is referred to as a *pre-coordinated concept* and can be referenced by its unique identifier. When a concept is represented using two or more concepts then it is referred to as a *post-coordinated concept* and this mechanism is known as *post-coordination*. Although existing SNOMED CT concepts can also be represented as post-coordinated concepts, the real power of post-coordination lies in its support in creation of new concepts which are not already present in SNOMED CT by identifying their relations with existing concepts. The concept of *severe pain in the stomach* which was shown in Figure 4 is a post-coordinated concept and is not already present in SNOMED CT. We defined it as a subtype of *pain (finding)* whose *finding-site* is *stomach structure* and whose *severity* is *severe* (the figure shows two ways of defining it as was explained earlier). We want to point out that any new post-coordinated concept will need to be well defined and hence it cannot be a primitive concept and can only be a fully-defined concept.

## 4. Material and Methods

Converting a clinical term into its SNOMED CT concept requires identifying all its related concepts. In this Section we describe our method to do this automatically. We first note that for primitive concepts the conversion is neither possible nor necessary. It is not possible because, for example, just from the term “asthma” itself, one cannot possibly infer that it is a disorder of respiratory system and that its finding site is airway structure, and on top of that these relations alone also do not fully define the concept of *asthma* which is a primitive concept. This conversion is also not necessary because primitive concepts are essentially basic concepts and are hence expressed in simple and short clinical terms that often exactly match one of the descriptions of the concept in SNOMED CT. For example, “asthma” will exactly match the fully-specified name of *asthma* thus rendering the conversion needless. We also note that any new clinical term mentioned in clinical text whose concept is not already in SNOMED CT cannot be a primitive concept. This is because if the new clinical term is expected to convey the idea of the clinical concept behind it then the concept must be fully expressed by the new clinical term, and thus it must be a fully-defined concept. For these reasons, our conversion method is only meant to work for fully-defined concepts.

### 4.1. Training for the Conversion Method

Our conversion method does not require any new manual annotations for training and learns only from clinical terms paired with their corresponding concepts already present in SNOMED CT. A big challenge for this conversion task is that there are over 300,000 concepts in SNOMED CT out of which the method needs to identify the correct related concepts for a given clinical term. As mentioned in the Related Work section, in our previous work [18], we had introduced the relation identification task, i.e. given two terms, classify whether they are related or not by a SNOMED CT relation. Although the kernel-based SVM classifier had worked well for the relation identification task, we found that it would not scale up for the full conversion task in which it needs to be applied to thousands of potential concepts in SNOMED CT in order to identify the related concepts. Hence in this work we present the following new and a much faster method that scales easily.

Our method first tokenizes clinical terms based on whitespaces and in the following we refer to tokens simply as words. Our method computes three kinds of probabilities from the training data through simple frequency counts. P($r$|$w$) is the probability that the relation $r$ will be present given that word $w$ is present in the term. For example, P(*is-a*(*disorder*,*disorder*) | "pain") is the probability that *is-a*(*disorder*,*disorder*) relation will be present in the concept given that the word "pain" is present in the clinical term. Note that this probability should be high because the word "pain" in a clinical term is a good indicator that the concept will be a type of a disorder.

Next, P($c$|$w$,$r$) is the probability that SNOMED CT concept $c$ (identified by its id) will be related by relation $r$ given that word $w$ is present in the term. For example, P(*stomach structure*|"gastric", *finding-site*(*disorder*,*body structure*)) is the probability that *stomach structure* will be the concept related by relation *finding-site*(*disorder*,*body structure*) given that word "gastric" is present in the clinical term. Please note that this relates words of clinical terms directly to the concepts. As was pointed out in the Related Work section, our previous method [18] was only relating words of the clinical term to the words of the related concept's description (e.g. "gastric" to "stomach" etc.) and thus only indirectly relating to the concept. It is helpful to relate words of the clinical terms directly to the concepts because the same words can appear in multiple concepts which can lead to errors.

However, it is still useful to also relate words of the clinical terms to words of the concepts' descriptions because in case there are not enough examples in the training data to learn P($c$|$w$,$r$) for a concept $c$ and word $w$, then the words present in the $c$'s description can also serve as indicators of the concept being related. To capture this, we compute a third kind of probability – P($v$|$w$,$r$) which is the probability that word $v$ will be present in the description of the concept which is related by relation $r$ given that word $w$ is present in the clinical term. For example, P("bacteria"|"bacterial*", causative-agent*(*disorder*,*organism*)) is the probability that the concept related by the relation *causative-agent*(*disorder*,*organism)* will have word "bacteria" present in its description given that the word "bacterial" is present in the clinical term.

Before estimating the probabilities from the training data, we first automatically normalize all the concepts in the training data as described in Section 3.3. Normalizing the concepts ensures that all the concepts are represented in a uniform way; for example, all the subtypes of *pain* will have *is-a* relation with *pain* and not with subtypes of *pain*, say, *abdominal pain* or *chest pain*. This uniformity aids in better training because it leads to more examples for learning the relations and the related concepts. We include only fully-defined concepts and their descriptions in the training data (of course their related concepts can be primitive) because we test the methods only on fully-defined concepts (as was explained earlier, the conversion is neither possible nor necessary for primitive concepts).

All the above three types of probabilities are estimated from the training data using simple frequency counts. P($r$|$w$) is estimated for each relation $r$ and each word $w$ from the training data by counting how many times the relation $r$ is present in the concept out of all the times a description of the concept had the word $w$ present in it. For implementation, we found it easier to first create, for every concept, a union of the words present in its descriptions including the fully-specified name. In the rest of the paper, when we say a word in a concept's description, we will refer to a word from this union. P($c$|$w$,$r$) is similarly estimated as the fraction of times concept $c$ is related by relation $r$ if the concept's description has word $w$ present and the concept has relation $r$ present. Finally, P($v$|$w$,$r$) is estimated as the fraction of times word $v$ is present in the description of the concept that is related by relation r if the concept's description has word $w$ present and the concept has relation $r$ present. The training only involves computing these frequency counts which is completed in just one pass over the training data and thus the method does not require much computational time even when it is trained on entire SNOMED CT.

### 4.2. Conversion Method

An overview of the conversion method is depicted in Figure 6 through an example. The example is explained later in this section after the description of the conversion method. For a given test clinical term $t$ with tokenized words $w_1, w_2, ..., w_n$, the conversion method proceeds as follows using the above computed probabilities. As the first step, the semantic type of its concept is determined. Out of all the *is-a(s,s)* relations, the relation which has the highest score for P(*is-a(s,s)*|$t$) is deemed to be the semantic type of the concept corresponding to $t$; this is depicted in the Equation 1 in Figure 5. P(*is-a(s,s)*|$t$), which is same as P(*is-a(s,s)*|$w_1, w_2, ..., w_n$), is computed by combining the individual P($r$|$w$) probabilities using the noisy-or method of combining probabilities [25], i.e. $P(r|w_1,w_2, ..., w_n) \approx 1-(1-P(r|w_1))*(1-P(r|w_2))*\ldots*(1-P(r|w_n))$, where $r$ stands for *is-a(s,s)*. Noisy-or is the probabilistic equivalent of logical *or* and is commonly used for combining probabilities in Bayesian networks [26,27]. If any of the individual probability scores is high, then the combined noisy-or probability score will be high. Furthermore, each individual probability score increases the combined probability score. This is exactly the behavior we want – presence of a single high probability word should be sufficient to indicate a relation and at the same time presence of multiple high probability words should further increase the probability of that relation. Another advantage of noisy-or combination is that presence of irrelevant words (i.e. those with probability P($r$|$w$) close to zero) does not negatively affect the combined probability (it is an “or” not an “and”). This is important because often in a clinical term there are words which are not indicative of presence of a particular relation but that should not decrease the probability of that relation. For example, presence of word “gastric” does not necessarily indicate presence of *is-a*(*disorder*,*disorder*) relation (it could be a *finding* or a *procedure* concept), however, that should not decrease the combined score for the presence of *is-a*(*disorder*,*disorder*) relation.

With the semantic type $s$ determined by the first step, the method next determines the related concepts. The semantic type of the concept restricts which relations could be present. Hence in this step only the relations that have domain semantic type of $s$ are considered, these will be of the form $k(s,u)$ where $k$ is the name of the relation and $u$ is the range semantic type. For example, if the semantic type $s$ is *disorder*, then the method would consider relations such as *finding-site*(*disorder*, *body structure*) and *causative-agent*(*disorder*, *substance*) among others. For all such relations $r=k(s,u)$, the related concept $c$ of semantic type $u$ is determined to be the one which has the maximum combinedScore($t,r,c$) as depicted in Equation 2 of Figure 5. This combined score (Equation 3) is a combination of three different scores (Equations 4-6) as explained below.

1) $$semantic\_type(t) = \underset{s}{\operatorname{argmax}}(P(is\_a(s,s)|t) \approx \underset{s}{\operatorname{argmax}}\left(1 - \prod_{i=1}^{n}\left(1 - P(is\_a(s,s)|w_i)\right)\right)$$

2) $$related\_concept(t,r) = \underset{c}{\operatorname{argmax}}\, combinedScore(t,r,c)$$

3) $$combinedScore(t,r,c) = ((score_1(t,c,r) + score_2(t,c,r) + 2 * score_3(t,c,r))/4) * P(r|t)$$

4) $$score_1(t,r,c) = P(c|w_1, w_2, \dots, w_n, r) \approx 1 - \prod_{i=1}^{n}(1 - P(c|w_i, r))$$

5) $$score_2(t,r,c) = P(v_1, v_2, \dots, v_m|w_1, w_2, \dots, w_n, r) \approx \frac{1}{m}\sum_{i=1}^{m} \max_{1 \le j \le n} P(v_i|w_j, r)$$

6) $$score_3(t,r,c) = similarity(\{v_1, v_2, \dots v_m\}, \{w_1, w_2, \dots, w_n\})$$

7) $$P(r|t) = P(r|w_1, w_2, \dots, w_n) \approx 1 - \prod_{i=1}^{n}(1 - P(r|w_i))$$

Figure 5. Equations to compute the semantic type and related concepts of a given clinical term *t* with words $w_1, w_2, \dots, w_n$. The semantic type *s* is first determined through Equation 1. Next, for every possible relation r=*k*(*s*,*u*) for that semantic type *s*, the best related concept *c* of semantic type *u* is determined using Equation 2 which utilizes Equations 3-7, where $v_1, v_2, \dots, v_m$ are words of concept *c*'s descriptions. The confidence score for the related concept is as computed using Equation 3.

The $score_1(t,r,c)$ is the probability $P(c|\, w_1, w_2, \dots, w_n, r)$, that is, the probability that concept *c* will be related by relation *r* if the clinical term has words $w_1, w_2, \dots, w_n$ present in it. This is computed using noisy-or probability combination over all the words $w_1, w_2, \dots, w_n$ (Equation 4). The $score_2(t,r,c)$ is the probability $P(v_1, v_2, \dots, v_m|w_1, w_2, \dots, w_n, r)$, that is, the probability that the concept related by relation r will have words $v_1, v_2, \dots, v_m$ in its description if the clinical term has words $w_1, w_2, \dots, w_n$ present in it. This probability is computed using individual $P(v_i|w_j)$ probabilities as the average of the maximum probability each word $v_i$ has among the words $w_j$ (Equation 5). We did it in this way instead of using a probability combination method such as noisy-or or multiplication because the number of words (*m*) in potentially related concepts' descriptions vary and hence noisy-or combination would mathematically favor concepts with more words in their descriptions while multiplication would mathematically favor concepts with fewer words in their descriptions. The above method of computing the probability is independent of the number of words in the potentially related concepts' descriptions and at the same time takes into account the best match each of these words has with the words in the given clinical term.

Often the similarity between two clinical terms is a good indicator of a potential relation between them, especially if less common words are present in the clinical term for which reliable estimates of probabilities could not be computed from the training data. Hence we also compute a third score, $score_3(t,r,c)$, which is a similarity score between the set of words $\{v_1, v_2, \dots, v_m\}$ and $\{w_1, w_2, \dots, w_n\}$ computed as the number of common words between them normalized to be between 0 and 1 (Equation 6). All the above three scores are then averaged. We found through a pilot study that giving twice the weight to the similarity score leads to slightly better results, hence we give twice the weight to the similarity score when averaging. The total score is divided by 4 so that the average will be between 0 and 1 (Equation 3). This averaged score represents the method's confidence in the concept *c* being related by relation *r*. But the method should also take into account the probability of the relation *r* being present because sometimes a clinical term itself indicates whether certain relations will be present in its concept or not. For example, a clinical term with "ulcer" is very likely to have a *finding-site*(*disorder*, *body structure*) relation but is less likely to have a *pathological-process*(*disorder*, *qualifier value*) relation. Hence the averaged score is also multiplied by the probability $P(r|t)$, that is, the probability of the relation *r* to be present given the clinical term *t*. This is computed using noisy-or combination of individual probabilities $P(r|w_i)$ (Equation 7). We call the resulting product as the combined score (Equation 3). Only the concept with the highest combined score is deemed to be related by the relation *k*(*s*,*u*) and that combined score is taken as the confidence of the method for that concept being related. During experimentation, if that confidence is above a chosen threshold then it is deemed that the concept is really related. The above process is repeated for each relation which has *s* as the domain semantic type. With different threshold values we can trade-off between precision and recall (the use of threshold is explained more in the next section).

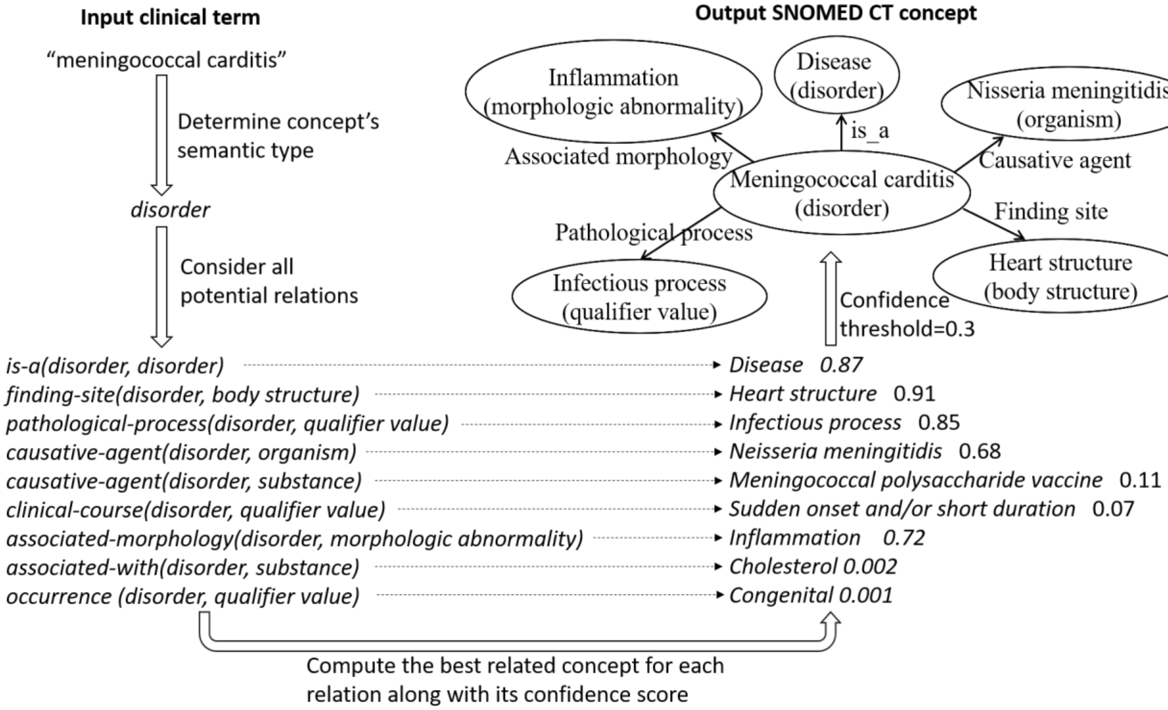

Figure 6. An illustration of how the method converts a given input clinical term into its SNOMED CT concept.

We restricted the method to only output one related concept for each relation because that is most common in SNOMED CT, especially after normalization. However, the method can be made to output more than one related concept if needed. Only primitive concepts are considered when finding the best related concept for an *is-a* relation because given that in the training data all the concepts were normalized, all the *is-a* relations were only with primitive concepts. Hence the method gets trained to identify related concepts for *is-a* relations only out of primitive concepts. For other relations, primitive as well as fully-defined concepts are considered as is the case with other relations in the training data. Please note that after normalizing a concept, its non-parent related concepts can be fully-defined concepts, although these concepts themselves upon their normalization will be represented using only primitive concepts.

Figure 6 illustrates the entire process of the method for a given input clinical term "meningococcal carditis". First, the method determines the most likely semantic type for its concept based on which *is-a* relation has the highest probability. In this case, *is-a*(*disorder*, *disorder*) had the highest probability hence the semantic type is determined as *disorder*. Next, the method considers all potential relations with domain semantic type of *disorder*. There were nine relations to consider in this case. For each relation, out of all the SNOMED CT concepts that have the range semantic type, the method determines the concept which is most likely to be related along with its confidence score. Finally, only the related concepts with confidence above the given threshold are included to be part of the output concept.

### 4.3. Evaluation Methodology

We did a large-scale automated evaluation of our method using existing concepts in SNOMED CT (2018 release) as well as did a small-scale manual evaluation using some clinical terms from real clinical text for which no concepts are present in SNOMED CT. For the first evaluation, we randomly selected 1000 concepts from SNOMED CT for testing and used the remaining concepts for training. All these test and training concepts were fully-defined because, as was explained earlier, the conversion is neither possible nor necessary for primitive concepts. As was also mentioned earlier, all concepts in SNOMED CT are marked to be either primitive or fully-defined. Then we repeated this process 10 times and macro-averaged all the evaluation measures over these 10 folds. The test concepts were out of four semantic types – *disorder*, *finding*, *procedure* and *pharmaceutical/biologic product*, which are the most important semantic types for conversion. Other semantic types such as, *organism*, *location* and *body structure*, generally do not need conversion because there are not many variations in their clinical terms and one rarely expresses a new concept of these semantic types in text. We considered 28 most frequent relations for our evaluation pertaining to the four concept semantic types that had at least one thousand instances in the entire SNOMED CT. Our method can be also applied to other relations, but they are too rare to be of significant importance or to affect our results.[2] For each test concept, one of its random description was used as input test clinical term. On average there were 3.43 relations per test concept. The output concepts are then compared with the normalized correct concepts. There is no need to normalize the output concepts because during testing, the related concepts for *is-a* relations are always chosen out of primitive concepts, and for other relations, as was pointed out earlier, normalization does not change the concepts that are related.

We measured the performance of our method on the conversion task in terms of *full conversion rate* which measures out of all the test clinical terms how many were converted fully into correct SNOMED CT concepts. Converted fully means that there was not a single mistake in identifying the relations and the related concepts for the 28 relations included in the evaluation. Please note that this is a very strict form of evaluation and gives no credit for even one extra relation or one missing relation or one incorrectly identified related concept. But in order to also gauge how far the method was from fully converting to the correct concepts when it did not, we also measured what we call *partial conversion rate* which allows for at most one extra, missing or incorrect relation. The evaluation was done on the combined output of the 10 folds. We varied the confidence threshold across the full range from 0 to 1 in intervals of 0.1 and report these two scores at the confidence threshold at which the method achieves the highest full conversion rate.

Besides measuring the conversion rates, we also measured performance of the method on the individual relations to see how well the method worked on identifying related concepts for each of those relations. For this evaluation, for each relation we generated precision-recall trade-off curve following the standard evaluation procedure [28]. Since we cannot show precision-recall for curves for 28 relations, we report only the maximum F-measures obtained over these curves. This evaluation was also done on the combined output of all the folds. For each relation, precision is – out of all the related concepts that the method identified how many were correct, and recall is – out of all the correct related concepts how many the method identified.

---

[2] We found that there were total 366 relations (each counted along with its domain and range) with one of the four semantic types as domain, however, many of these had very few instances, for example, 240 of them had less than 10 instances and 128 of them had only 1 instance; rarer relations seem to be errors in SNOMED CT. Hence we decided to limit the number of relations and excluded those with less than 1000 instances. We also excluded relations involving numerical types (for example, the quantity of pharmaceutical/biologic product) because they are easy to handle separately using a numeric type. The included relations covered 90% of the instances.

Finally, we also did a small-scale evaluation on clinical terms found in real clinical texts. We used a corpus that was recently created for the different task of clinical term normalization [5]. In that corpus, clinical terms found in clinical text were annotated with their respective concepts in SNOMED CT. However, if a clinical term was present in text whose corresponding concept was not in SNOMED CT then it was left untagged. But in such cases, as was mentioned earlier, if the clinical term had subterms whose concepts were present in SNOMED CT then those subterms were annotated with those corresponding concepts. For example, if the text had the term “left breast biopsy” whose concept is not present in SNOMED CT, then the annotators instead annotated “left” and “breast biopsy” with their respective SNOMED CT concepts. We automatically gathered all such cases in which two annotated clinical terms were adjacent in the corpus because that is an indication of a larger clinical term for which there is no concept in SNOMED CT. We randomly selected 35 such terms for which we were able to manually create post-coordinated concepts (note that it is not possible to create SNOMED CT concepts for all clinical terms because SNOMED CT has its limitations [29,30]). The number of words in these clinical terms varied from 2 to 5 with an average of 3.06 words. After training it on the entire SNOMED CT, we evaluated our method on this small corpus by measuring the full and the partial conversion rates in the same way as for the large-scale evaluation. This small-scale evaluation dataset is available on the website http://www.uwm.edu/~katerj/conversion.

## 5. Results and Discussion

We first evaluated our method using the existing SNOMED CT terms and concepts according to the methodology described in the last section. Table 1 shows the results for both full conversion and partial conversion rates. In the first row, the performance of the complete method is shown that must determine the semantic type of the given clinical term (first step of the method described in Section 4.2). In the second row, we show the performance when the correct semantic type is given to the method. Having the knowledge of correct semantic types for clinical terms is realistic for some applications, for example, if the clinical terms have been extracted automatically using an information extraction system [31] for various entity types such as diseases, procedures, etc., then the semantic type of the term will be already known. However, the gain in performance when correct semantic types are given is only marginal, this is because the method is already very good at automatically determining the semantic types. We found that it achieved 95.4% accuracy on the task of determining the semantic types of clinical terms. It can be seen from the table that the method is able to fully convert nearly one-third of clinical terms without making a single mistake in identifying their relations and the related concepts. If we allow at most one mistake, then the method can convert nearly two-third of clinical terms.

**Table 1**.Performance of the method at converting clinical terms into concepts in SNOMED CT’s ontology evaluated using existing SNOMED CT terms and concepts. Full conversion means there is not a single mistake in identifying all the related concepts while partial conversion means at most one mistake is overlooked.

| Method | Full conversion rate (%) | Partial conversion rate (%) |
|---|---|---|
| Complete method | 32.2 | 62.54 |
| With semantic type given | 33.23 | 64.68 |

In order to determine if there are differences between the method’s performance for different relations, we also show evaluation results for each relation. Table 2 shows the maximum F-measure along with the corresponding precision and recall for each relation. Please note that this evaluation is very different from the evaluation presented in our past work [18] in which we had shown the performance on the relation identification task, that is, given two clinical terms, whether their concepts are related by a relation or not. In contrast, here we are showing results of identifying the correct related concept for a relation (or its absence) out of all the thousands of possible concepts in SNOMED CT. This is clearly different and a much harder task. Table 2 also shows the number of instances (n) of each relation in the test data of total 10,000 examples. The number of instances ranged from the highest of 5,181 instances for the *is-a(disorder*, *disorder*) relation to the lowest of 110 instances for the *using-access-device*(*procedure*, *physical object*) relation.

**Table 2.** Maximum F-measures (F) and the corresponding precisions (P) and recalls (R) in percentages for the twenty-eight relations. The number of instances (n) of each relation in the test data of total 10,000 examples are also shown.

| Relation | n | F | P | R | Relation | n | F | P | R |
|---|---|---|---|---|---|---|---|---|---|
| is-a(pharmaceutical/biologic product, pharmaceutical/biologic product) | 1030 | 91.97 | 98.61 | 86.21 | using-access-device(procedure, physical object) | 110 | 75.51 | 79.89 | 72.4 |
| has-manufactured-dose-form (product, dose form) | 588 | 89.79 | 93.04 | 86.87 | is-a(finding, finding) | 1189 | 74.85 | 83.4 | 67.94 |
| associated-with(disorder, substance) | 321 | 89.52 | 85.78 | 93.9 | has-basis-of-strength-substance (pharmaceutical/biologic product, substance) | 430 | 71.57 | 73.69 | 70.76 |
| is-a(procedure, procedure) | 2928 | 89.24 | 92.32 | 86.37 | has-active-ingredient (pharmaceutical/biologic product, substance) | 1134 | 71.56 | 82.02 | 63.55 |
| occurrence (disorder, qualifier value) | 418 | 84.99 | 88.69 | 82.1 | using-device(procedure, physical object) | 320 | 71.02 | 78.07 | 65.85 |
| is-a(disorder, disorder) | 5181 | 84.11 | 86.26 | 82.07 | associated-morphology (disorder, morphologic abnormality) | 3587 | 69.19 | 73.15 | 65.71 |
| clinical-course(disorder, qualifier value) | 262 | 83.52 | 82.12 | 85.48 | procedure-device(procedure, physical object) | 190 | 67.18 | 69.19 | 66.47 |
| has-specimen(procedure, specimen) | 190 | 82.86 | 86.41 | 80.37 | using-substance(procedure, substance) | 148 | 64.86 | 72.98 | 61.89 |
| causative-agent(disorder, substance) | 972 | 81.74 | 87.38 | 76.87 | causative-agent(disorder, organism) | 588 | 63.23 | 74.21 | 55.4 |
| pathological-process(disorder, qualifier value) | 811 | 81.32 | 83.52 | 79.48 | direct-substance(procedure, substance) | 308 | 62.79 | 74.96 | 55.53 |
| has-intent(procedure, qualifier value) | 210 | 78.6 | 96.21 | 67.34 | procedure-site(procedure, body structure) | 2468 | 62.17 | 69.99 | 55.95 |
| method(procedure, qualifier value) | 2996 | 77.67 | 83.93 | 72.31 | interprets(finding, observable entity) | 1244 | 61.55 | 83.99 | 48.72 |
| has-interpretation(finding, qualifier value) | 757 | 76.93 | 88.49 | 68.3 | finding-site(disorder, body structure) | 4431 | 54.29 | 60.19 | 49.58 |
| direct-morphology(procedure, morphologic abnormality) | 510 | 75.57 | 81.36 | 70.82 | finding-site(finding, body structure) | 480 | 51.82 | 69.52 | 41.84 |

The results shown in Table 2 are in the sorted order of maximum F-measure. As can be seen, the method worked well for many relations and obtained maximum F-measures in eighties and seventies. But the method obtained low performance on a few relations. We found that there were two major types of errors. The first type of error was when the method would identify a closely related concept instead of the exact concept; for example, it would identify *ankle joint structure* instead of *ankle region structure* as the *finding-site*, or it would identify *occipital lobe structure* instead of *occipital bone structure*. The second type of error was due to missing a finer and a rarer concept for a broader and a more common concept; for example, the method would identify *lower limb structure* instead of *peripheral nerve of lower limb* as the *finding-site*. This type of error is most likely because the method sees more training examples that include common concepts and not many that include finer and rarer concepts. These two types of errors were especially common for the related concepts of *body structure* semantic type. It can be also observed from Table 2 that the two relations on which the method performed lowest had *body structure* as the range semantic type, and one of them had 4,431 instances in the test data of total 10,000 examples. These errors adversely affected our method's full conversion rate because in computing the full conversion rate no credit is given even when there is only one error in conversion.

We want to point out that for certain applications, having only partial conversion can still be helpful. For example, if the method is being used to help a human to identify or create new SNOMED CT concepts then the human in the loop can benefit from partial conversion who can then complete or correct the rest of the conversion Furthermore, the methods developed in this work for determining the semantic type and related concepts can potentially also assist in auditing ontologies [32]. For example, they can be useful for auditing semantic type assignments [33, 34], for auditing or finding missing *is-a* [35,36] as well as other types of relations [37].

**Table 3.** Results of the ablation study in which $score_1$, $score_2$ and $score_3$ components of the combined probability scores (Figure 5) are respectively excluded individually. The table shows Maximum F-measure averaged across all relations. Results of the baseline that uses only similarity scores is shown in the last row.

| Method | Averaged Max F-measure (%) |
|---|---|
| Complete method | 73.74 |
| Without $score_1$ | 37.23 |
| Without $score_2$ | 47.83 |
| Without $score_3$ | 69.3 |
| Baseline: Using only similarity scores ($score_3$) | 15.25 |

We also did an ablation study to determine the contribution of various components of our method. Table 3 shows the results. For each method, we computed the average of the maximum F-measures of all the relations weighted by the number of correct concepts related by each relation (shown as n in Table 2). The complete method (corresponding to Table 2) obtained average maximum F-measure of 73.74%. Excluding any of the three types of scores as described in Section 4.2 and shown in Figure 5 adversely affected the performance. However, the reduction in performance was least when not using $score_3$ (similarity score) showing that it is the least important component and similarity score does not help as much as the probability scores. The reduction was drastic for the other two scores and was highest for $score_1$. This shows the value of probabilities $P(c|\ w, r)$, that is, relating words of the clinical terms directly to SNOMED CT concepts as was pointed out earlier in the Related Work section. The last row of Table 3 shows the performance of a baseline in which only the similarity scores are used to determine relations. It obtained a very poor performance thus showing that just similarity between clinical terms is not sufficient to determine related concepts.

**Table 4.** Performance of the method at converting clinical terms into post-coordinated concepts in SNOMED CT's ontology evaluated using clinical terms from real clinical text for which corresponding concepts did not exist in SNOMED CT.

| Method | Full conversion rate (%) | Partial conversion rate (%) |
|---|---|---|
| Complete method | 25.71 | 54.29 |
| With semantic type given | 28.57 | 62.86 |

Besides evaluating our method on a large-scale using existing SNOMED CT concepts, we also evaluated it on a small-scale using terms from real clinical texts whose corresponding concepts were not present in SNOMED CT as described in Section 4.3. For this evaluation, entire SNOMED CT was used for training. The results are shown in Table 4 for the complete method that must determine the semantic type (first row) and when the correct semantic type is given to the method (second row). Although lower, the numbers are not widely different from what we had obtained when evaluated on SNOMED CT terms and concepts (Table 1). Misidentifying the *body structure* was again a major source of errors similar to what had happened when evaluated on SNOMED CT. Another source of errors in this evaluation was misidentifying the semantic type of the concept. The accuracy of identifying the semantic type was 80% in this evaluation which is much less than the corresponding accuracy when evaluated on SNOMED CT (95.4%). Hence the gain in performance is higher in this evaluation when the correct semantic types are given. This evaluation was on a small-scale hence it is far from being conclusive, but the results generally indicate that the evaluation on SNOMED CT was representative of how the method would perform on clinical terms from real clinical text.

## 6. Limitations and Future Work

Our method was based on learning word-level probabilities from training data and did not use any word sequence, syntactic, or other linguistic information. Although that made our method very fast and easily scalable, a possibility of improvement in the future will be to leverage more information from the structure of clinical terms by first linguistically analyzing them. In future, incorporating word similarity through means such as word embeddings may lead to further improvement. Another possibility of future work is to utilize the ontological structure of SNOMED CT for better conversion. Currently our method treats all SNOMED CT concepts as separate and does not exploit connections between them. It should be helpful, for example, to utilize the hierarchical structure of concepts in SNOMED CT so that when fewer examples are available to learn probabilities for a finer concept, back-off probabilities from its parent or ancestor concepts could be utilized.

The current work focused on converting a clinical term into its concept in SNOMED CT by finding all the related concepts that defines it. However, it did not consider nested relations in which the method may need to find further relations of the related concepts, for example, a disorder caused by a pharmaceutical/biologic product which in turn will have its own related concepts. SNOMED CT allows such nesting or compositional definition of concepts. Although such concepts are less likely to be encountered in clinical text, in future, it will be interesting to extend the current method to be able to do this. One possibility will be to apply the conversion method recursively according to the compositional structure of the clinical term. In this work, the form of the evaluation was hard, that is, no credit was given if the method did not find the exact related concept. In future, one could come up with a softer form of evaluation in which partial credit will be given if the method finds a semantically close concept instead of the exact concept, for instance, if the method finds the parent or a subtype concept. It will, of course, depend on the application task whether finding a semantically close concept instead of the exact concept will be still useful or will be harmful, and hence whether a soft form of evaluation will be viable or not for that application.

Besides evaluating our method on existing SNOMED CT terms and concepts on a large-scale, we also evaluated it on a small-scale on a manually annotated corpus of clinical terms found in real clinical text. However, to be more representative, this evaluation needs to be done on a larger scale which will require creating a large manually annotated corpus of clinical terms and their post-coordinated concepts. We point out that it is not easy to create such a dataset because ideally it requires annotators to have both clinical and ontological expertise. It will be also an interesting avenue for future work to explore whether the current method itself can help in creating such a dataset by first converting clinical terms to their concepts which can then be verified or corrected by human experts.

## 7. Conclusions

We presented, to the best of our knowledge, the first complete method to convert clinical terms into their concepts in SNOMED CT's ontology, including to concepts which are not already listed in SNOMED CT. For a given clinical term, the method converts it into a SNOMED CT concept by identifying the related concepts in SNOMED CT that define the concept. The method is based on computing probabilities from the training data and is computationally simple, fast and scalable. It does not require any additional manual annotation and learns only from existing SNOMED CT terms paired with their concepts. The method performed well in identifying related concepts for most of the relations. But there is scope for improving the full conversion rate which depends on identifying every single related concept correctly. Leveraging linguistic structure of clinical terms, ontological structure of SNOMED CT, compositionality, and word similarity are possible avenues for further improvement.

## References

1 SNOMED International. http://www.snomed.org/ (Accessed June 16, 2020).

2 Baader, Franz, Diego Calvanese, Deborah McGuinness, Peter Patel-Schneider, and Daniele Nardi, eds. The description logic handbook: Theory, implementation and applications. Cambridge university press, 2003.

3 Wu, Stephen T., Vinod C. Kaggal, Dmitriy Dligach, James J. Masanz, Pei Chen, Lee Becker, Wendy W. Chapman, Guergana K. Savova, Hongfang Liu, and Christopher G. Chute. "A common type system for clinical natural language processing." Journal of biomedical semantics 4, no. 1 (2013): 1.

4 Pradhan, Sameer, Wendy Chapman, Suresh Man, and Guergana Savova. "Semeval-2014 task 7: Analysis of clinical text." In Proc. of the 8th International Workshop on Semantic Evaluation (SemEval) (2014): 54–62.

5 Luo, Yen-Fu, Weiyi Sun, and Anna Rumshisky. "MCN: A comprehensive corpus for medical concept normalization." Journal of biomedical informatics (2019): 103132.

6 Jung, Sungwon, Seunghee Kim, Sooyoung Yoo, and Jinwook Choi. "Toward the automatic generation of the entry level CDA documents." *Journal of Korean Society of Medical Informatics* 15, no. 1 (2009): 141-151.

7 Barrett, Neil, Jens Weber-Jahnke, and Vincent Thai. "Automated clinical coding using semantic atoms and topology." In *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 1-6. IEEE, 2012.

8 Stenzhorn, Holger, Edson Jose Pacheco, Percy Nohama, and Stefan Schulz. "Automatic mapping of clinical documentation to SNOMED CT." *Stud Health Technol Inform*. 150 (2009): 228-232.

9 Ruch, P., Gobeill, J., Lovis, C. and Geissbühler, A., 2008, October. Automatic medical encoding with SNOMED categories. In BMC medical informatics and decision making (Vol. 8, No. 1, p. S6). BioMed Central.

10 Leaman, Robert, Rezarta Islamaj Doğan, and Zhiyong Lu. "DNorm: disease name normalization with pairwise learning to rank." Bioinformatics 29, no. 22 (2013): 2909-2917.

11 Pradhan, Sameer, Noemie Elhadad, Brett R. South, David Martinez, Lee M. Christensen, Amy Vogel, Hanna Suominen, Wendy W. Chapman, and Guergana K. Savova. "Task 1: ShARe/CLEF eHealth Evaluation Lab 2013." In *CLEF (Working Notes)*. 2013.

12 National NLP Clinical Challenges (n2c2), Track 3: n2c2/UMass Track on Clinical Concept Normalization. https://n2c2.dbmi.hms.harvard.edu/track3 (Accessed June 16, 2020).

13 Kate, Rohit J. "Normalizing clinical terms using learned edit distance patterns." *Journal of the American Medical Informatics Association* 23, no. 2 (2016): 380-386.

14 Luo, Yen-Fu, Weiyi Sun, and Anna Rumshisky. "A Hybrid Normalization Method for Medical Concepts in Clinical Narrative using Semantic Matching." AMIA Summits on Translational Science Proceedings 2019 (2019): 732-740.

15 Li, Haodi, Qingcai Chen, Buzhou Tang, Xiaolong Wang, Hua Xu, Baohua Wang, and Dong Huang. "CNN-based ranking for biomedical entity normalization." BMC bioinformatics 18, no. 11 (2017): 79-86.

16 Petrova, Alina, Yue Ma, George Tsatsaronis, Maria Kissa, Felix Distel, Franz Baader, and Michael Schroeder. "Formalizing biomedical concepts from textual definitions." *Journal of biomedical semantics* 6, no. 1 (2015): 22.

17 Miñarro-Giménez, Jose Antonio, Catalina Martínez-Costa, Pablo López-García, and Stefan Schulz. "Building SNOMED CT post-coordinated expressions from annotation groups." Stud Health Technol Inform 235 (2017): 446-450.

18 Kate, Rohit J. "Towards converting clinical phrases into SNOMED CT expressions." Biomedical informatics insights 6 (2013): BII-S11645.

19 Cristianini, Nello, and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

20 Peterson, Kevin J., and Hongfang Liu. "Automating the Transformation of Free-Text Clinical Problems into SNOMED CT Expressions." *AMIA Summits on Translational Science Proceedings* 2020 (2020): 497-506.

21 Aronson, Alan R. "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program." In Proceedings of the AMIA Symposium, American Medical Informatics Association (2001):17-21.

22 SNOMED CT Terminology Services Guide. https://confluence.ihtsdotools.org/display/DOCTSG/ (Accessed June 16, 2020).

23 Bhattacharya, S. B. "Introduction to SNOMED CT." (2016) Springer.

24 SNOMED CT Editorial Guide. https://confluence.ihtsdotools.org/display/DOCEG (Accessed June 16, 2020).

25 Zagorecki, A., & Druzdzel, M. (2004). An empirical study of probability elicitation under noisy-OR assumption. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2004*, *2*, 880–885.

26 Oniśko, Agnieszka, Marek J. Druzdzel, and Hanna Wasyluk. "Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates." *International Journal of Approximate Reasoning* 27, no. 2 (2001): 165-182.

27 Vomlel, Jiří. "Noisy-or classifier." *International Journal of Intelligent Systems* 21, no. 3 (2006): 381-398.

28 Japkowicz, Nathalie, and Mohak Shah. *Evaluating learning algorithms: a classification perspective*. Cambridge University Press, 2011.

29 Lee, Dennis H., Francis Y. Lau, and Hue Quan. "A method for encoding clinical datasets with SNOMED CT." *BMC medical informatics and decision making* 10, no. 1 (2010): 53.

30 Liu, Hongfang, Kavishwar Wagholikar, and Stephen Tze-Inn Wu. "Using SNOMED-CT to encode summary level data–a corpus analysis." AMIA Summits on Translational Science Proceedings 2012 (2012): 30-37.

31 Wang, Yanshan, Liwei Wang, Majid Rastegar-Mojarad, Sungrim Moon, Feichen Shen, Naveed Afzal, Sijia Liu et al. "Clinical information extraction applications: a literature review." *Journal of biomedical informatics* 77 (2018): 34-49.

32 Zheng, Ling, Zhe He, Duo Wei, Vipina Keloth, Jung-Wei Fan, Luke Lindemann, Xinxin Zhu, James J. Cimino, and Yehoshua Perl. "A review of auditing techniques for the Unified Medical Language System." *Journal of the American Medical Informatics Association* (2020).

33 Fan, Jung-Wei, and Carol Friedman. "Semantic reclassification of the UMLS concepts." *Bioinformatics* 24, no. 17 (2008): 1971-1973.

34 He, Zhe, Yehoshua Perl, Gai Elhanan, Yan Chen, James Geller, and Jiang Bian. "Auditing the assignments of top-level semantic types in the UMLS semantic network to UMLS concepts." In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1262-1269. IEEE, 2017.

---

35 Cui, Licong. "COHeRE: Cross-ontology hierarchical relation examination for ontology quality assurance." In *AMIA Annual Symposium Proceedings*, vol. 2015, pp. 456-465. American Medical Informatics Association, 2015.

36 Chen, Yan, Huanying Helen Gu, Yehoshua Perl, and James Geller. "Structural group-based auditing of missing hierarchical relationships in UMLS." *Journal of biomedical informatics* 42, no. 3 (2009): 452-467.

37 Vizenor, Lowell T., Olivier Bodenreider, and Alexa T. McCray. "Auditing associative relations across two knowledge sources." *Journal of biomedical informatics* 42, no. 3 (2009): 426-439.