Transitioning from Text-Based to Code-Based NLP Rules using SNOMED CT for Data Normalization and Standardization

Michael Denton MS, RN, Jacee Robison MS, RN, Jonathan Herbert, RN, Rachael Howe, MS, RN
3M Health Information Systems, Murray, Utah

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Introduction

A pilot project was completed to evaluate the feasibility of using SNOMED CT concepts to discover clinically relevant information from free-text narratives through an existing natural language processing (NLP) application using hard-coded text strings.

Background

The legacy NLP application required users to manually add hard-coded text strings to the NLP rules to ensure that every variation of a clinical concept would be found in free-text narratives. The process for adding and maintaining text strings was resource intensive and inefficient due to the lack of a standardized, concept-based model.

Purpose

- Improve the accuracy and amount of discovered clinically relevant information by implementing a terminology model to normalize and standardize hard-coded text strings.
- Map legacy hard-coded text strings to equivalent pre-coordinated SNOMED CT concepts. Where pre-coordinated concepts do not exist, apply mapping heuristics to determine a set of post-coordinated terms to implement into the terminology model.
- Establish a process for terminology scaling and maintenance.

Image 1: *NLP discovery of hard-coded text strings*

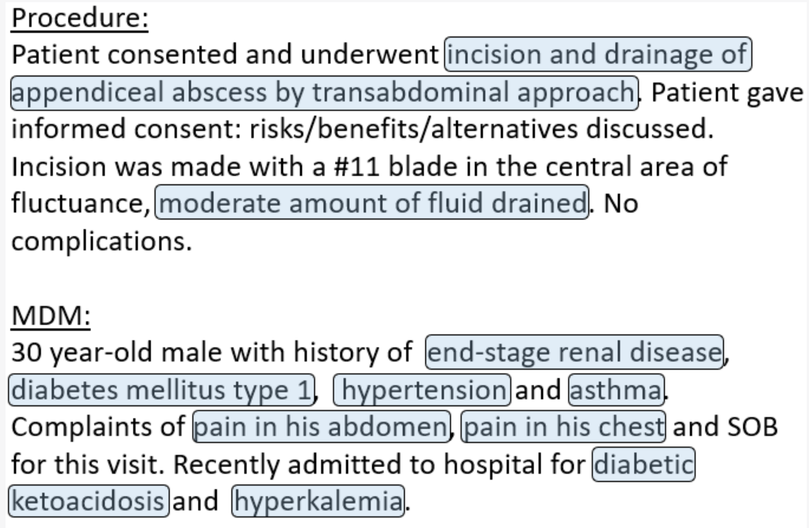

Table 1: *Examples of legacy hard-coded text strings*

Text Strings
Incision and drainage of appendiceal abscess by transabdominal approach
Moderate amount of fluid drained
End-stage renal disease
Diabetes mellitus type 1
Hypertension
Asthma
Pain in his abdomen
Pain in his chest
Diabetic Ketoacidosis
Hyperkalemia

Transitioning from Text-Based to Code-Based NLP Rules using SNOMED CT for Data Normalization and Standardization

Michael Denton MS, RN, Jacee Robison MS, RN, Jonathan Herbert, RN, Rachael Howe, MS, RN
3M Health Information Systems, Murray, Utah

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Methods

The project methodology consisted of three phases:

1. Evaluation
2. Mapping
3. Modeling

Evaluation:

Included the analysis of 11,106 text strings for clinical terms and the identification of patterns for mapping heuristics. Heuristics were used by subject matter experts (SME) to determine the level of specificity required for a clinically equivalent concept when mapping to SNOMED CT.

Mapping:

Involved the creation of a mapset between legacy hard-coded text strings and SNOMED CT concepts with applied mapping heuristics.

Modeling:

Consisted of creating pre-coordinated concepts by combining post-coordinated terms to cover clinically equivalent concepts identified in text strings for the terminology model.

Image 2: *Text String Mapping to One-to-Many Concepts*

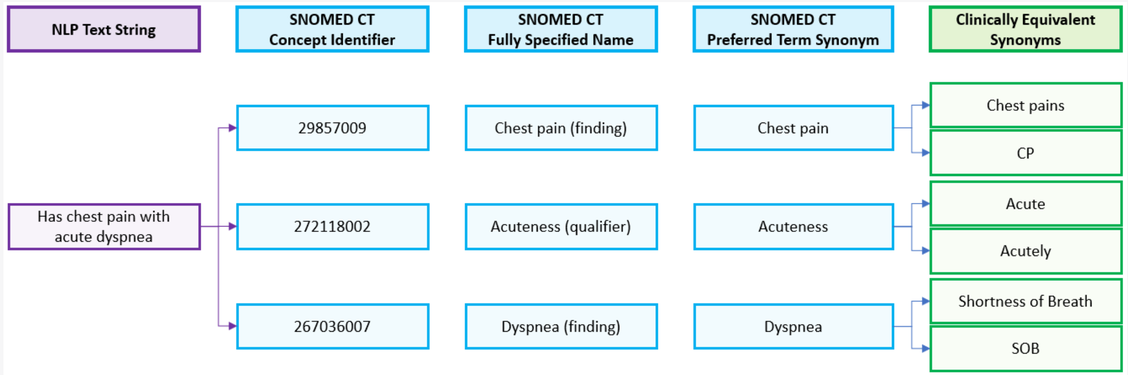

Image 2 displays an example of the mapping process from legacy hard-coded text strings to SNOMED CT concepts. Using mapping heuristics, SME determined that the following text string "*Has chest pain with acute dyspnea*" suggested a one-to-many mapping to clinical terms "*chest pain*" and "*acute dyspnea*". SNOMED CT does not have the pre-coordinated concept "acute dyspnea" but does have post-coordinated terms "acuteness" and "dyspnea". By using a terminology model, we were able to combine post-coordinated terms into a pre-coordinated concept and add clinically equivalent synonyms.

Transitioning from Text-Based to Code-Based NLP Rules using SNOMED CT for Data Normalization and Standardization

Michael Denton MS, RN, Jacee Robison MS, RN, Jonathan Herbert, RN, Rachael Howe, MS, RN
3M Health Information Systems, Murray, Utah

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Results

The project resulted in 11,106 text strings being mapped to 6,657 unique concepts using SNOMED CT pre and post-coordinated terms for the suggestion of medical diagnoses, diagnostic tests, and supporting clinical information.

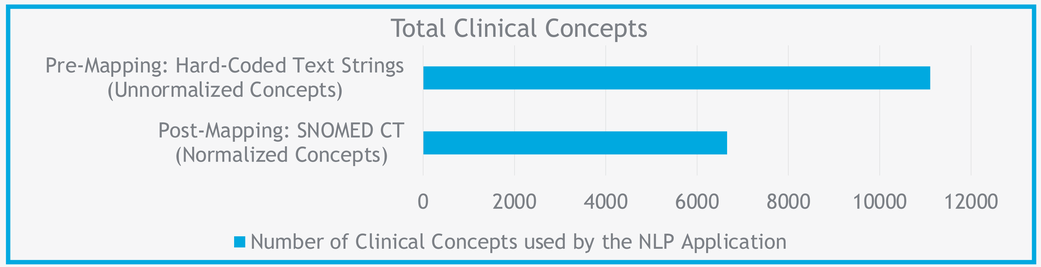

Table 2: *Translation of Text Strings (Image 1) to SNOMED CT Concepts (Image 3)*

SNOMED CT Identifier (SCTID)	Fully Specified Name (FSN)
37402004	Incision and drainage of abscess of appendix (procedure)
6736007 33463005 122462000	Moderate (severity modifier) (qualifier value) Liquid substance (substance) Drainage procedure (procedure)
392521001 46177005	History of (contextual qualifier) (qualifier value) End-stage renal disease (disorder)
472970003	History of diabetes mellitus type 1 (situation)
161501007	History of hypertension (situation)
161527007	History of asthma (situation)
21522001	Abdominal pain (finding)
29857009	Chest pain (finding)
267036007	Dyspnea (finding)
420422005	Ketoacidosis due to diabetes mellitus (disorder)
14140009	Hyperkalemia (disorder)

Image 3: *NLP discovery of SNOMED CT Concepts*

Procedure:
Patient consented and underwent incision and drainage of appendiceal abscess by transabdominal approach. Patient gave informed consent: risks/benefits/alternatives discussed. Incision was made with a #11 blade in the central area of fluctuance, moderate amount of fluid drained. No complications.

MDM:
30 year-old male with history of end-stage renal disease, diabetes mellitus type 1, hypertension and asthma. Complaints of pain in his abdomen, pain in his chest and SOB for this visit. Recently admitted to hospital for diabetic ketoacidosis and hyperkalemia.

Transitioning from Text-Based to Code-Based NLP Rules using SNOMED CT for Data Normalization and Standardization

Michael Denton MS, RN, Jacee Robison MS, RN, Jonathan Herbert, RN, Rachael Howe, MS, RN
3M Health Information Systems, Murray, Utah

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Discussion

The NLP application required us to use a terminology model to handle concepts in both a pre and post-coordinated format. Where pre-coordinated concepts did not exist, post-coordinated concepts were combined in the terminology model to be used by the NLP. As shown in *Image 2*, there is not a pre-coordinated concept in SNOMED CT for the entire text string "*Has chest pain with acute dyspnea*". To accommodate this in the NLP, the terminology model allowed us to combine three clinical concepts from SNOMED CT ("*chest pain*", "*acute*", and "*dyspnea*") to suggest the same clinically equivalent text string without having to hard-code each unique text string variation.

Conclusions

The project results demonstrate a successful translation from hard-coded text strings to one-to-one or one-to-many SNOMED CT concepts (*Table 2*). By translating hard-coded text strings to codable concepts we were able to utilize a terminology model to normalize and standardize clinical concepts for the NLP application to discover. In addition, the concept-based approach allowed users to create an efficient and generalizable process for improved concept scaling and maintenance.

Lessons Learned and Future Directions

The translation from hard-coded text strings to SNOMED CT concepts demonstrated comprehensive coverage of clinical content. However, one of the limitations discovered in the project was the need for additional clinical content structured in both pre and post-coordinated formats.

Future work will include creating a larger terminology model library through a SNOMED CT extension to support a broader set of customer use cases with SNOMED CT being the foundation of the terminology model.

In addition, the following considerations will need to be evaluated in order to scale clinical content discovery through continuous quality improvement:

- Data driven testing
- Content models & templates
- Bulk content creation
- Narrative contexts – temporality, experiencer, assertion & negation
- Cross terminology mapping