

Representing Clinical Diagnostic Criteria in Quality Data Model Using Natural Language Processing

Na Hong

Mayo Clinic/ Rochester, MN,
USA

Institute of Medical Information,
Chinese Academy of Medical Sciences/
Beijing, China

Hong.na@mayo.edu

Dingcheng Li

Mayo Clinic/ Rochester,
MN, USA

Li.dingcheng@mayo.edu

Yue Yu

Mayo Clinic/ Rochester, MN,
USA

School of Public Health, Jilin
University/ Changchun, China

Yu.yue@mayo.edu

Hongfang Liu

Mayo Clinic/ Rochester, MN,
USA

Liu.Hongfang@mayo.edu

Christopher G. Chute

Johns Hopkins University/
Baltimore, MD, USA

chute@jhu.edu

Guoqian Jiang

Mayo Clinic/ Rochester, MN,
USA

Jiang.guoqian@mayo.edu

Abstract

Constructing standard and computable clinical diagnostic criteria is an important and challenging research area in clinical informatics community. In this study, we present our framework and methods for representing clinical diagnostic criteria in Quality Data Model (QDM) using natural language processing (NLP) technologies. We used a clinical NLP tool known as cTAKES for preprocessing of textual diagnostic criteria. We created mappings between cTAKES type system and QDM elements in both datatype and data levels. We evaluated the performance of our NLP-based approach by annotating 218 individual diagnostic criteria in the categories of Symptom and Laboratory Test. In conclusion, our NLP-based approach is a feasible solution in developing diagnostic criteria representation and computerization.

1 Introduction

The term *diagnostic criteria* designates the specific combination of signs, symptoms, and test results that the clinician uses to attempt to determine the correct diagnosis¹. It is one kind of the most valuable sources of knowledge for supporting clinical decision-making and improving pa-

tient care (Yager and McIntyre, 2014). Diagnostic criteria are a critical evidence resource of clinical decision support system; however, diagnostic criteria are usually described without uniform standard, scattered over different media such as medical textbooks, literatures and clinical practice guidelines, and mostly in free text formats. Several methods based on natural language processing (NLP) technology have been reported and used in structuring free-text-based clinical guidelines, clinical notes and electronic health records (EHRs), as (Rea, etc., 2012) and (Ohno-Machado, etc., 2013). However, there are not sufficient researches on using NLP-based approaches to support the formalization of free-text diagnostic criteria. To achieve computable diagnostic criteria, we consider that a computable model to represent diagnosis criteria and the use of clinical NLP applications to support the modeling are two essential research areas.

Current efforts on development of international recommendation standard models in clinical domains have laid the foundation for modeling and representing computable diagnostic criteria. National Quality Forum (NQF) Quality Data Model (QDM) (Quality Data Model, 2015) as an information model that describes clinical concepts in a standardized format. It allows quality measure developers and many clinical researchers or performers to describe clearly and unambiguously the data required to calculate the performance measure. QDM is designed with the purpose to

¹
https://en.wikipedia.org/wiki/Medical_diagnosis#Diagnostic_criteria

allowing EHRs (Li, et al., 2012) and other clinical electronic systems to share a common understanding and interpretation of the clinical data. In a previous study, researcher Jiang (2015) evaluated the application feasibility of QDM through a data-driven approach and demonstrated that the use of QDM is feasible in building a standards-based information model for representing computable diagnostic criteria.

On clinical NLP studies, many NLP tools currently are applied in the clinical unstructured free text processing and also support terminology annotation, such as Health Information Text Extraction tool (HITex)², MetaMap (Aronson and Lang, 2010), OpenNLP³ and Clinical Text Analysis and Knowledge Extraction System (cTAKES) (Savova, et al., 2010). Some studies compared the performance of the frequently used NLP tools, and the results showed that cTAKES scored best in both performance and usability. cTAKES is an open source Apache project and it is a NLP system for extraction of information from electronic medical record clinical free-text. cTAKES was built on the Unstructured Information Management Architecture (UIMA) framework which is an open source framework designed by IBM and a series of comprehensive NLP methods (Bruce, 2012). In this study, we use cTAKES as a NLP tool to support the formalization of diagnostic criteria.

The objective of our study is to describe our efforts in developing a semi-automatic approach using NLP to facilitate the representation of clinical diagnostic criteria in QDM.

2 Materials & Methods

2.1 Materials

cTAKES: The components of cTAKES are specifically trained for the clinical domain, and create rich linguistic and semantic annotations that can be utilized by clinical decision support systems and clinical research⁴. cTAKES discovers clinical named entities and clinical events using a dictionary lookup algorithm and a subset of the Unified Medical Language System (UMLS)⁵, mainly including the following mentions: disease/disorders, sign/symptoms, medications, anatomical sites and procedures.

Besides, cTAKES extract named entity attributes and assigns values for the attributes such as UMLS concept unique identifiers (CUIs) and SNOMED CT codes, polarity, uncertainty, conditional, etc. In this study, we used the cTAKES version 3.2.1.

NQF QDM: The QDM consists of criteria for data elements, relationships for relating data element criteria to each other, and functions for filtering criteria to the subset of data elements that are of interest⁶. The basic components of the QDM include: category (e.g., Symptom), datatype (e.g., Symptom, Active), attribute (e.g., information about severity, start Datetime, stop Datetime, and ordinality), and value set comprising concept codes from one or more code systems. In this study, we used the QDM version 4.1 (Quality Data Model, 2015).

2.2 Methods

Figure 1 shows a framework we designed for the NLP-supported QDM modeling of diagnostic criteria. The framework comprises three modules. The first module is an NLP annotation module. We use cTAKES as a NLP tool to support structured representation of diagnostic criteria. The second module is a data model transformation between cTAKES type system and QDM elements. The transformation is supported using both manual mapping strategies and machine learning algorithms. The third module is a unified web interface for human review. As the output, all collected data elements, value sets and logic expressions of diagnostic criteria are formalized by using QDM-based standard representation.

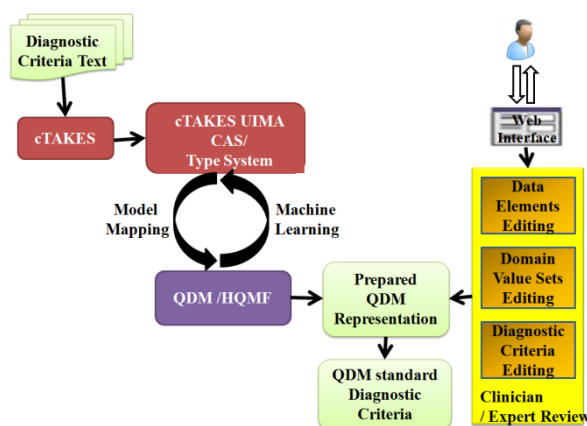


Figure 1. A framework for the NLP-supported QDM modeling of diagnostic criteria.

²https://www.i2b2.org/software/projects/hitex/hitex_manual.html.

³<https://opennlp.apache.org/>

⁴<http://en.wikipedia.org/wiki/CTAKES>

⁵<http://www.nlm.nih.gov/research/umls/>

⁶<http://www.healthit.gov/quality-data-model>.

2.2.1 NLP annotation

We first used the cTAKES to perform NLP annotation on textual diagnostic criteria. cTAKES is a modular system of pipelined components combining rule-based and machine learning techniques, introduced in (Savova, Masanz, etc., 2010). As an operable interface, UIMA provides the tooling for selecting which descriptors are used together and determining the order of the descriptors, see detail in (cTAKES 3.2 Component Use Guide, 2015). Dictionaries such as UMLS, SNOMED CT and RxNorm are integrated into cTAKES clinical pipeline.

2.2.2 Data Model Transformation

We implemented the model mapping and data transformation on two levels: the datatype-level mapping and the data-level mapping.

(1) Datatype-level Mapping

We created the datatype-level mappings between cTAKES UIMA Common Analysis System (CAS) type system and QDM datatypes, as well as corresponding attributes and features between these two heterogeneous schemas (Figure 2). We established the mapping relations through analyzing the textual definitions of datatypes in both models. Datatype-level mappings are mainly focused on 7 selected QDM datatypes and 8 cTAKES types that frequently appear in diagnostic criteria.

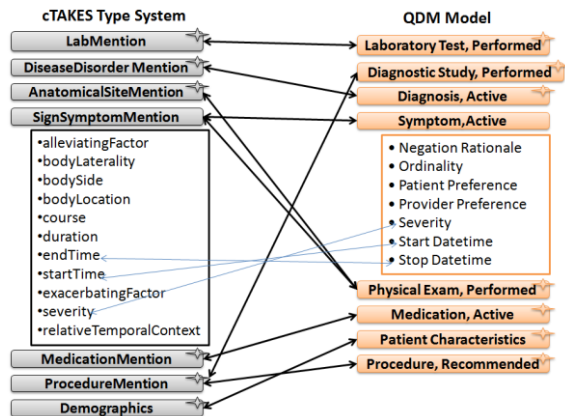


Figure 2. Datatype-level Mapping between cTAKES type system and QDM elements

(2) Data-level Mapping

The second level of mapping analysis is the data-level mappings which are created between the data structure of cTAKES CAS and QDM Health Quality Measure Format (HQMF) (HQMF Templates for QDM December 2013, 2015). The cTAKES processes text and stores the results in the UIMA-CAS structure, whereas the HQMF as a standard format is used to represent QDM-

based eMeasure data. All cTAKES instance data output as CAS XML data and are converted into HQMF XML data using the data-level mapping rules.

Figure 3 illustrates the data-level mapping rules between CAS and HQMF elements we created for QDM datatype Laboratory Test, Performed.

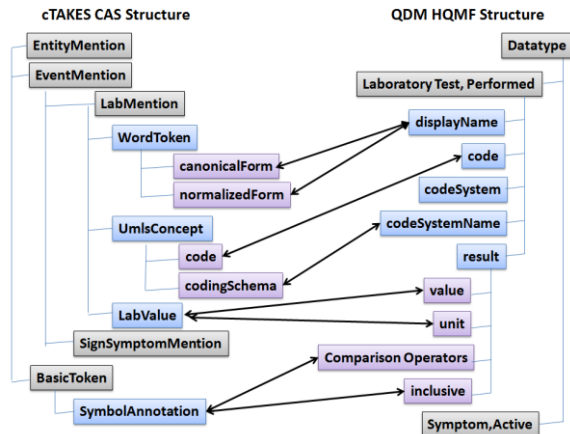


Figure 3. Data-level mappings between cTAKES CAS XML and QDM HQMF XML

2.2.3 Evaluation

For evaluation, we manually annotated a collection of individual criteria with QDM datatypes and attributes. We used the manual annotations as gold standard and evaluated the performance of NLP-based annotations. Two authors (HN, GJ) reviewed the annotations and the consensus was resolved through discussions. Three standard measures were used to describe the performance of the NLP module: precision, recall and F-measure.

3 Results

To implement experiment and evaluation, we first collected 218 individual criteria in the Symptom and Laboratory Test categories. The individual criteria were extracted manually from the text of 44 diagnostic criteria in 13 different clinical topics (an example of textual diagnostic criteria is shown in Appendix A). All the diagnostic criteria are collected from a number of sources including medical textbooks, journal papers, documents issued by professional organization (such as the World Health Organization - WHO) and Internet. Table 1 shows the number of individual criteria of the 13 clinical topics. We used a cTAKES (V.3.2.1) NLP analysis engine known as the *AggregatePlaintxtUMLSProcessor* and processed the test criteria. Using the datatype-

level mapping rules we created and the cTAKES annotation results of two EventMentions (*LabMention* and *SignSymptomMention*), our algorithm automatically allocated a QDM datatype for each individual diagnostic criteria.

Table 1 Clinical Topics Distribution of 218 individual criteria.

Clinical Topic	Number of Symptom Criteria	Number of laboratory test Criteria
Allergy	6	0
Cardiology	45	15
Critical Care	37	26
Dermatology	18	5
Diabetes	2	9
Endocrinology and Metabolism	27	5
Gastroenterology	1	2
Hematology	0	4
Immunology	2	3
Infectious Disease	0	4
Nephrology	0	4
Neurology	1	1
orthopedics	1	0
Total	140	78

The allocation results could reflect the automatic mapping classification performance for QDM datatypes (*Laboratory Test, Performed* and *Symptom, Active*). For example, Figure 4 and Figure 5 show the text of two individual diagnostic criteria with CTAKES annotations in *LabMention* and *SignSymptomMention*.

Example1: *Thrombocytopenia (platelets <100,000 cells/mm3)*

Thrombocytopenia (platelets <100,000 cells/mm3)

Figure 4. “platelets” is highlighted as an cTAKES annotation in *LabMention*

Example2: *Gastrointestinal-hepatic dysfunction: Moderate (diarrhea, nausea/vomiting, abdominal pain)*

Gastrointestinal hepatic dysfunction:
Moderate (diarrhea, nausea/vomiting, abdominal pain)

Figure 5. “diarrhea, nausea, vomiting, abdominal pain” are highlighted as the cTAKES annotations in *SignSymptomMention*.

After automatically mapping based on our mapping rules between two systems, diagnostic criteria in free-text are transformed into QDM based HQML XML structure. One of the QDM data examples attached in the Appendix B. Table 2 shows the evaluation results in terms of whether mapping rules correctly allocate a QDM datatype for an individual criterion.

To evaluate the performance of data-level mappings, we tested the mapping results of 78 individual diagnostic criteria which were annotated manually using the *QDM datatype Laboratory Test, Performed*. The test was mainly focused on the mapping performance of four attribute elements, including code/code system, laboratory test value, measurement and unit, comparison operator. Table 3 shows the evaluation results of elements mapping in the QDM datatype *Laboratory Test, Performed*.

Table 2. Performance of the Datatype-level Mapping Results

QDM Datatype	Laboratory Test, Performed	Symptom, Active
Precision	94.0%	69.2%
Recall	80.8%	59.3%
F-score	86.9%	63.9%

Table 3. Performance of the Data-level Elements Mapping Results (QDM: Laboratory Test, Performed)

QDM Element	Code/Code System	Value	Unit	Operator
Precision	94.0%	96.3%	100%	61.9%
Recall	80.8%	98.2%	53%	26.5%
F-score	86.9%	97.2%	69.3%	37.1%

4 Discussion

To bridge the semantic gap between cTAKES type system and QDM Model, we performed critical element analysis and created element mappings in both datatype and data levels. As cTAKES UIMA-CAS and QDM both are comprehensive models with independent structures, more semantic analysis need to be studied in order to extend our current mapping rules, e.g., the mapping analysis on QDM temporal representation and cTAKES temporal type. Furthermore, there exist elements that could not be directly mapped between two models under different contexts.

Previous studies investigated the eligibility criteria in clinical trial protocol and developed approaches (known as EliXR) for eligibility criteria extraction and semantic representation, and used hierarchical clustering for dynamic categorization of such criteria (Weng , etc., 2011) (Luo, etc.,2011). In future, we will develop machine learning-based methods leveraging the EliXR approach to enable the analysis for a large amount of clinical diagnostic criteria data.

The study demonstrated overall performance of cTAKES used for generating the QDM-based representation of diagnostic criteria. The evaluation results in Table 2 indicated that criteria in the Laboratory Test category could be automatically classified into the QDM datatype effectively; whereas the performance for classifying criteria in the Symptom category was sub-optimal. The reason is mainly because that cTAKES uses the *SignSymptomMention* that doesn't distinguish between a sign and a symptom. The evaluation results in Table 3 indicated that the code/code system and value mappings could acquire satisfactory performance whereas the performance for the unit annotation is good in precision but sub-optimal in recall. In addition, the operator recognition was insufficient, for examples, in criteria 'Sézary cells with a diameter > 14 um representing > 20% of the circulating lymphocytes', '%' is annotated as Symbol but '>' is not recognized in cTAKES that cause low precision. Above all, the mapping rules were able to generate validated QDM datatypes and related elements, covering most typical model elements used in diagnostic criteria. NLP-based technologies could provide a semi-automatic way to support the preliminary classification and enable a pattern-based QDM representation.

5 Conclusion

In this study, we demonstrated that clinical NLP tool (e.g., cTAKES) could support the QDM modeling of free-text diagnostic criteria in a semi-automatic way. We are actively working on developing machine learning algorithms to improve the performance of our NLP-based approaches for representing clinical diagnostic criteria in QDM.

Reference

Joel Yager, John S. McIntyre. 2014. *DSM-5 Clinical and Public Health Committee: Challenges and Considerations*, American Journal of Psychiatry, 171: 142-44.

Susan Rea, Jyotishman Pathak, Guergana Savova, Thomas A. Oniki, Les Westberg, Calvin E. Beebe, Cui Tao, Craig G. Parker, Peter J. Haug, Stanley M. Huff, Christopher G. Chute. 2012. *Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPN project*, Journal of Biomedical Informatics, 45: 763-71.

Lucila Ohno-Machado, Prakash Nadkarni, Kevin Johnson. 2013. *Natural language processing: algorithms and tools to extract computable information from EHRs and from the biomedical literature*, Journal of the American Medical Informatics Association, 20: 805-05.

Alan R Aronson , François-Michel Lang. 2010. *An overview of MetaMap: historical perspective and recent advances*, Journal of the American Medical Informatics Association, 17: 229-36.

Quality Data Model. [cited May 10, 2015]. Available from: <http://www.healthit.gov/quality-data-model>.

Guoqian Jiang, Harold R. Solbrig, Jyotishman Pathak, Christopher G. Chute. 2015. *Developing a Standards-based Information Model for Representing Computable Diagnostic Criteria: A Feasibility Study of the NQF Quality Data Model*, MedInfo (in submission)

Guergana K. Savova, James J. Masanz, Philip V. Ogren , Jiaping Zheng , Sunghwan Sohn , Karin C. Kipper-Schuler , Christopher G. Chute. 2010. *Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications*, J Am Med Inform Assoc, 17: 507-13.

Lars-Erik Bruce. 2012. *Apache UIMA and Mayo cTAKES UIMA and how it is used in the clinical domain*. [cited May 10, 2015]: Available from: <http://www.uio.no/studier/emner/matnat/ifi/INF5880/v12/undervisningsmateriale/seminar.pdf>

cTAKES 3.2 Component Use Guide. [cited May 10, 2015]. Available from: <https://cwiki.apache.org/confluence/display/CTAKES/cTAKES+3.2+Component+Use+Guide>.

National Quality Forum, HQMF. 2013. *Templates for QDM December*. [cited May 10, 2015]: Available from: http://www.healthit.gov/sites/default/files/qdm_hqmf_templates_dec2013.pdf

Chunhua Weng , Xiaoying Wu , Zhihui Luo , Mary Regina Boland , Dimitri Theodoratos , Stephen B Johnson. 2011. *EliXR: an approach to eligibility criteria extraction and representation*, J Am Med Inform Assoc, 18 Suppl 1: i116-24.

Zhihui Luo, Meliha Yetisgen-Yildiz, Chunhua Weng. 2011. *Dynamic categorization of clinical research eligibility criteria by hierarchical clustering*, J Biomed Inform, 44: 927-35.

Dingcheng Li, Cory M Endle, Sahana Murthy, Craig Stancl, Dale Suesse, Davide Sottara, Stanley M. Huff, Christopher G. Chute, Jyotishman Pathak. 2012. *Modeling and executing electronic health records driven phenotyping algorithms using the NQF Quality Data Model and JBoss(R) Drools Engine*, AMIA Annu Symp Proc, 2012: 532-41.

Appendix

(A) An example of textual diagnostic criteria for diabetes mellitus

1) $A1C \geq 6.5\%$. The test should be performed in a laboratory using a method that is NGSP certified and standardized to the DCCT assay.

OR

2) $FPG \geq 126$ mg/dl (7.0 mmol/l). Fasting is defined as no caloric intake for at least 8 h.

OR

3) 2-h plasma glucose ≥ 200 mg/dl (11.1mmol/l) during an OGTT. The test should be performed as described by the World Health Organization, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.

OR

4) In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥ 200 mg/dl (11.1 mmol/l).

In the absence of unequivocal hyperglycemia, criteria 1–3 should be confirmed by repeat testing.

Reference: American Diabetes Association.2012. *Diagnosis and Classification of Diabetes Mellitus*, Diabetes Care. 35(suppl_1):S64-S71.

(B) An QDM representation of diagnostic criterion based on cTAKES annotation

Example text: *Thrombocytopenia (platelets <100,000 cells/mm3)*

```
<sourceOf typeCode="PRCN">
  <conjunctionCode code="AND"/>
  <act classCode="ACT" moodCode="EVN" isCriterionInd="true"><!-- Laboratory Test, Result pattern -->
    <templateId root="2.16.840.1.113883.3.560.1.12"/>
    <id root="c5244e91-3c2e-4863-ae87-a48556b9e3ae"/>
    <code code="30954-2" displayName="Results" codeSystem="2.16.840.1.113883.6.1"/>
    <sourceOf typeCode="COMP">
      <observation classCode="OBS" moodCode="EVN" isCriterionInd="true">
        <code code="2.16.840.1.113883.3.117.1.7.1.267" displayName="Platelets Count LOINC Value Set"
          codeSystem="2.16.840.1.113883.3.560.101.1"/>
        <title>Laboratory Test, Result: platelets (result < 100,000 cells/mm3)</title>
        <statusCode code="completed"/>
        <sourceOf typeCode="REFR">
          <observation classCode="OBS" moodCode="EVN" isCriterionInd="true">
            <templateId root="2.16.840.1.113883.3.560.1.1019.3"/>
            <code code="385676005" codeSystem="2.16.840.1.113883.6.96" displayName="result"
              codeSystemName="SNOMED-CT"/>
            <value xsi:type="IVL_PQ">
              <high value="100,000" unit="cells/mm3" inclusive="false"/>
            </value>
          </observation>
        </sourceOf>
      </observation>
    </sourceOf>
  </act>
</sourceOf>
```