MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190553

Top-Level Design of a Normalized Chinese Clinical Terminology: An Integrated Application of National and International Data Standards and Terminologies

Haixia Sun[a], Yujing Ji[a], Panpan Deng [a], Junlian Li [a], Huiling Ren[a], Liu Shen[a], Ming Feng[b], Yi Wang[b]

[a] Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China
[b] Peking Union Medical College Hospital, Beijing, China

Abstract

This work describes the design and building of a Chinese clinical terminology (called CCTS). The terminology is similar to an ontology, and will promote the use of Chinese clinical data, such as indexing, retrieval and exchange. The terminology is a TOPL concept framework, which integrates hierarchical structures of Chinese and interenational reference terminology standards for health. Our framework includes 14 subtrees, 2286 classes and 65 relationships.

Keywords:
Vocabulary, Electronic Health Records, Reference Standards

Introduction

Wide implementation of electronic health record systems in China has made clinical data available electronically. However, little effort has been devoted to building a national clinical terminolgy, limiting the use of clinical data for indexing, retrieval, and exchange. Motivated by this, we developed a Chinese clinical terminology, named *Chinese Clinical Terminology System (CCTS)*. We divided this work into two phases: the first phase is the top-level design, constructing the top concept framework of CCTS; the second phase is the refinement and expansion, collection and organization of more terms and concepts according the top concept framework. In this paper we introduce the top-level design.

Methods

First, we analyzed main knowledge types in Electronic Medical Records (EMR) according to the Specification for Sharing Document of Electronic Medical Record and Basic Dataset of Electronic Medical Record released by National Health Commission of the People's Republic of China (NHCPRC). This analysis indicated the scope of the terminology. Domain experts then manually selected the 14 top classes.

Then, we extracted the top 5 level concepts from terminologies and code systems (Table 1). These concepts are recommended in Chinese and international health information standards. The 5 level concepts were extracted in the following steps:

1. Manually determine the best reference subtree for different classes by evaluating and ranking them for different classes
2. Merge other related subtrees by string matching and rules
3. Recommend preferred term for classes according to reference books published by the China National Committee for Terms in Sciences and Technologies (CNCTST)
4. Inspect hierarchical relationship manually
5. Define other medical semantic relationships among the first 5 level classes referring to semantic relationships in the Unified Medical Language System (UMLS)

Results

Our top-level design phase produced 2369 high level concept classes, organized into 14 subtrees (Table 2) with 65 kinds of semantic relationships. We recommend 136 core reference sources for lower subtrees. We are in the process of extracting and fusing lower subtrees from those sources.

Discussion

This paper describes the development of CCTS through a top-level concept framework that integrates hierarchical structures of Chinese and international reference standards for health terminology, as well as expert knowledge. Reference sources for each subtree to build lower subtrees were also recommended, which will be used to guide the construction of the terminology system.

Information model design for different concept classes and extracting Out-of-Vocabulary words from EMRs using medical natural language processing (MedNLP) technology will be used in future work. We anticipate that CCTS will cover common medical terms and concepts in various EMR systems in China, and can be used to greatly facilitate EMR searching, retrieval, clustering, and reasoning, by providing rich sets of synonyms and various clinical relationships. Correspondingly, it also can promote the interoperability of EMRs by intergating Chinese and international health reference standards.

Conclusions

Design of the top concept framework is the first step towards building a domain terminology system or ontology. Integrating multiple resources is an efficient way to achieve this goal, and domain experts are indispensable. We envision the top concept framework will be used to guide the next phase of CCTS.

Table 1– Some Core References for The 14 Subtrees

Top Classes	Core References
疾病诊断 (Disease Diagnosis)	Disease Classification and Code (GB/T14396, China); ICD 10, SNOMED CT, Disease Ontology
有机体(Organism)	MeSH
解剖部位(Anatomy)	FMA, MeSH
诊查对象(Observable and Examinable Object)	SNOMED CT
临床表现 (Clinical Manifestation)	Health Information Data Metadata Catalogue Part 6: Chief Complaints and Symptoms (WS 363.6, , China); SNOMED CT, HPO
诊疗项目、技术和方法 (Diagnosis and Treatment Item, Technique and Method)	Clinical test item classification and code (WS/T102, China), Health information data element catalogue Part 9: Laboratory examination (WS 363.9, China), ICD-9-CM-3, LOINC
化学药品和生物制品 (Chemical Drug and Biological Product)	ATC, Chinese Pharmacopoeia 2015
医用设备、器械和材料 (Medical Equipment, Instrument and Material)	Social Insurance Medical Service Classification and Code (LD/T01-2017, China), Health information data element value code Part 16: Drugs, equipment and materials (WS 364.16, China)
物质(Substance)	SNOMED CT
心理行为(Psychology and Behavior)	SNOMED CT
环境地理 (Environment and Geography)	World and Region Name Codes (GB/T2659-2000, China)
事件、事故和灾害 (Event, Accident and Disaster)	ICD-10, MeSH, SNOMED CT
人口学及社会经济学特征 (Demographic and Socioeconomic Characteristic)	Health Information Data Metadata Catalogue Part 3: Demographic and socioeconomic characteristics (WS 363.3-2011, China), Health information data element value code Part 3: Demographic and socioeconomic characteristics (WS 364,3-2011, China)
限定语(Qualifier)	SNOMED CT

Table 2– The Number of First 5 Level Classes

Subtrees	Level 2	Level 3	Level 4	Level 5
Disease Diagnosis	24	145	434	183
Organism	4	13	83	9
Anatomy	18	78	45	6
Observable and Examinable Object	27	-	-	-
Clinical Manifestation	5	38	54	1
Diagnosis and Treatment Item, Technique and Method	14	86	258	232
Chemical Drug and Biological Product	25	98	98	-
Medical Equipment, Instrument and Material	11	40	149	-
Substance	9	36	26	7
Psychology and Behavior	3			
Environment and Geography	7	4	10	2
Event, Accident and Disaster	3	25	-	-
Demographic and Socioeconomic Characteristic	2	27	6	-
Qualifier	24	-	-	-

Acknowledgements

This paper was funded by The Chinese Academy of Medical Sciences (Grant No. 2017-I2M-3-014)

References

[1] Fisher H M, Hoehndorf R, Bazelato B S, et al., DermO; an ontology for the description of dermatologic disease, *Journal of Biomedical Semantics 7* (2016), 38-46.

[2] Dramé K, Diallo G, Delva F, Dartigues JF, Mouillet E, Salamon R, Mougin F, Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: an application to Alzheimer's disease. *J Biomed Inform* 48 (2014), 171-82.

[3] Malhotra A, Younesi E, Gündel, Michaela, et al., ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease, *Alzheimer's and Dementia* 10.2 (2014), 238-246.

Address for correspondence
Corresponding author : Haixia Sun
Email: sun.haixia@imicams.ac.cn