The Value of SNOMED CT Description for the Mapping of Chinese Rare Diseases Ontology

Zhi Wang| Digital China Health

SNOMED CT Expo 2019
Kuala Lumpur | Oct 31-Nov 1

E-POSTERS
SPONSORED BY:
tpp

Click on a bubble to navigate the e-poster

introduction
methods
results
discussion

Abstract

Rare diseases (RDs) are rare, but rare disease patients are numerous. More and more interests are paid on RDs in China. A list containing 121 rare diseases (Rare Diseases List, RDL) was published by five governmental departments of China, including National Health Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology, National Medical Products Administration, State Administration of Traditional Chinese Medicine, in June of 2018. National Rare Diseases Registry System, containing 128 rare diseases now, is a platform for the data registry of rare diseases in China. Since China is not a member country of SNOMED CT, NRDRS obtained a first affiliate license of SNOMED CT in China. The author and his team were authorized to use this license to build up an ontology named CRDO, Chinese Rare Diseases Ontology, based on the 208 rare diseases from RDL and NRDRS.

One of the use cases of CRDO is knowledge linking. The Each concept in CRDO has an English name, which was used for linkage with Radiology Gamuts Ontology (RGO) that is a domain application ontology for radiology and a knowledge resource for radiology diagnosis. Since the concepts were mapping with SNOMED CT, which contained a large number of synonyms in SNOMED CT description, the synonyms in SNOMED CT were also used for augmenting the matching between CRDO and RGO for those that don't have an exact match (indirect exact match) between concepts in CRDO and RGO.

There are 92 CRDO concepts exactly mapping with RGO, in which 77 were direct and 15 indirect. The method was also used in the mapping between CRDO and other terminologies, including MeSH, ICD-10, ICD-11 and OMIM. There were 66 direct and 12 indirect for mapping to MeSH, 9 direct and none indirect for ICD-10, 59 direct and 3 indirect for ICD-11 and 49 direct and 12 indirect for OMIM.

The plenty of synonyms in SNOMED CT is a wonderful bridge between terminologies.

Background

According to the definition of United States and the countries of European Union, the diseases with the patient's population of 0.065%~0.1% is called rare diseases (RDs). Despite the low prevalence of single diseases, the overall affected population can represent a striking proportion due to the great variety of rare diseases, thus contributing to substantially high disease burden for patients, families, and the community all around the world. Considering the vast population base of China, the disease burden can add up to an astronomical figure. The world's largest rare disease database has identified nearly 7000 diseases, but only about 400 rare diseases have drugs that are available and often cost prohibitively. Due to the small number of RD patients, low market demand, and high costs of research and development, few pharmaceutical companies care about the development of RD treatment drugs, so these drugs are aptly called "orphan drugs."

A list containing 121 rare diseases (Rare Diseases List, RDL) was published by five departments of China, including National Health Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology, National Medical Products Administration, State Administration of Traditional Chinese Medicine, in June of 2018. The National Rare Diseases Registry System of China (NRDRS) is a project based on the National Key Research and Development Program for Precision Medicine, the "Rare Disease Clinical Cohort Study" project (project number: 2016YFC0901500), which was constructed by Peking Union Medical College Hospital under the help of Digital China Health (http://www.dchealth.com).

The Value of SNOMED CT Description for the Mapping of Chinese Rare Diseases Ontology

Zhi Wang| Digital China Health

SNOMED CT Expo 2019
Kuala Lumpur | Oct 31-Nov 1

E-POSTERS
SPONSORED BY:
tpp

Click on a bubble to navigate the e-poster

introduction
methods
results
discussion

Materials and Methods

NRDRS has the first affiliate license and cooperates with Orphanet. The author and his team learned and passed the e-learning courses provided by SNOMED International. They used the mechanisms and methods of SNOMED CT and ORDO to build up CRDO, which followed the logic model and concept model of SNOMED CT and the hierarchy of ORDO.

CRDO Components

CRDO served as the main terminology for NRDRS and related medical terms. Figure 1 showed the CRDO design. Concepts, descriptions and relationships were the three types of components, which represented the contents of CRDO.

- Concepts represents the clinical meanings;
- Descriptions links the human readable terms to concepts;

Relationships links each concept to others.

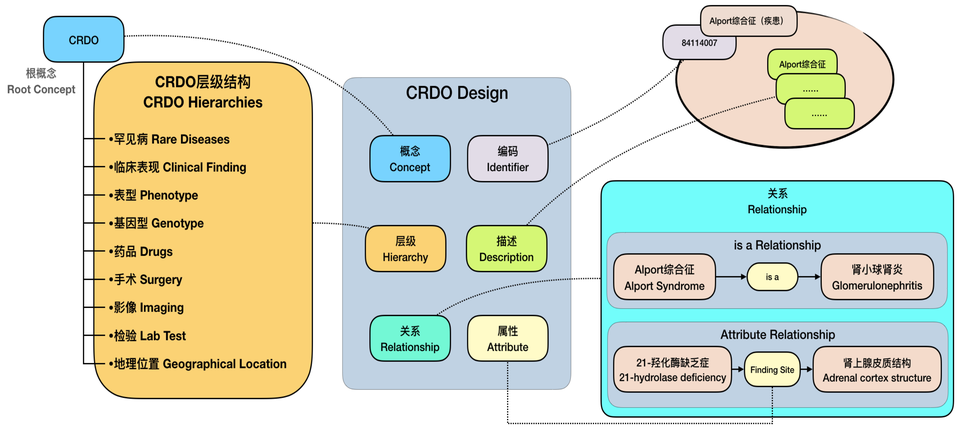

Figure 1 CRDO Design

Concepts and Hierarchies

CRDO concepts represented the clinical idea related to rare diseases, including the diseases, clinical findings, phenotypes, genotypes, drugs, surgery, imaging, laboratory tests and geographical location which served as the top-level hierarchies.

The hierarchy of RDs was the core element of CRDO. The 121 diseases from RDL were one of the basic inputs of RDs, part of which was showed in Table 1. The other input was from the disease names from NRDRS. All the diseases were merged to form the main hierarchy of CRDO. There were 203 concepts for rare diseases in CRDO.

Table 1 Part of the diseases of RDL.

	Disease Name (Chinese)	Disease Name (English)
1	21-羟化酶缺乏症	21-hydroxylase deficiency
2	白化病	Albinism
3	Alport 综合征	Alport Syndrome
4	肌萎缩侧索硬化	Amyotrophic Lateral Sclerosis
5	Angelman 氏症候群(天使综合征)	Angelman Syndrome
6	精氨酸酶缺乏症	Arginase Deficiency
7	热纳综合征(窒息性胸腔失养症)	Asphyxiating Thoracic Dystrophy(Jeune Syndrome)
8	非典型溶血性尿毒症	Atypical Hemolytic Uremic Syndrome
9	自身免疫性脑炎	Autoimmune Encephalitis
10	自身免疫性垂体炎	autoimmune hypophysitis
11	自身免疫性胰岛素受体病	Autoimmune Insulin Receptopathy(Type B insulin resistance)
12	β-酮硫解酶缺乏症	Beta-ketothiolase Deficiency
13	生物素酶缺乏症	Biotinidase Deficiency
14	心脏离子通道病	Cardiac Ion Channelopathies
15	原发性肉碱缺乏症	Carnitine Deficiency
16	Castleman 病	Castleman Disease
17	腓骨肌萎缩症	Charcot-Marie-Tooth Disease
18	瓜氨酸血症	Citrullinemia

The Value of SNOMED CT Description for the Mapping of Chinese Rare Diseases Ontology

Zhi Wang| Digital China Health

SNOMED CT Expo 2019
Kuala Lumpur | Oct 31-Nov 1

E-POSTERS
SPONSORED BY:
tpp

Click on a bubble to navigate the e-poster

introduction

methods

results

discussion

Result

In CRDO Version 1.0, there are 1920 concepts with 7677 axioms. By combining the diseases names from RDL and the ones from NRDRS, there are 203 concepts representing rare diseases in CRDO.

There were six types for all the mappings for CRDO mapped to SNOMED CT, ORDO, RGO, MeSH, ICD-11, OMIM and ICD-10, shown in Table 2, which showed an example of the mappings for one of the RDs, named 21-hydroxylase deficiency. Table 3 showed the “Indirect matches” between CRDO, SNOMED CT and RGO.

We built up a web page for CRDO 1.0, which was published to NRDRS website (www.nrdrs.org.cn), in which the conditions that can result in or from the rare diseases and the references information, which were from RGO, and the cross mappings were shown, see Figure 2.

Figure 2 CRDO web page in NRDRS.

Table 2 The result of terminology mapping of CRDO.

Target Terminology	Exact Match		Supertype Mapping		No Match	Not Sure
	Direct	Indirect	Direct	Indirect		
SNOMED CT	161	3	0	4	35	0
ORDO	149	0	0	0	54	0
RGO	77	15	0	0	111	0
MeSH	66	12	0	1	118	6
ICD 11	59	3	84	3	54	0
OMIM	49	12	0	0	98	44
ICD 10	9	0	131	0	63	0

Table 3 “Indirect matches” between CRDO, SNOMED CT and RGO.

CRDO ID	CRDO Disease Name (Chinese)	CRDO Disease Name (English)	SNOMED CT ID	SNOMED CT Concept Name	RGO ID	RGO Concept Name
1448005	多种酰基辅酶A脱氢酶缺乏症	Multiple Acyl-CoA Dehydrogenase Deficiency	22886006	Glutaric aciduria, type 2 (disorder)	23624	Glutaric aciduria type 2
3473004	庞贝氏病	Pompe disease	274864009	Glycogen storage disease, type II (disorder)	4574	Glycogen storage disease type 2
3471002	库欣综合征	Cushing’s syndrome	47270006	Hypercortisolism (disorder)	33962	Hypercortisolism
1626008	脊髓延髓肌萎缩症(肯尼迪病)	Spinal and Bulbar Muscular Atrophy (Kennedy Disease)	5262007	Spinal muscular atrophy (disorder)	29953	Spinal muscular atrophy

The Value of SNOMED CT Description for the Mapping of Chinese Rare Diseases Ontology

Zhi Wang| Digital China Health

SNOMED CT Expo 2019
Kuala Lumpur | Oct 31-Nov 1

E-POSTERS
SPONSORED BY:
tpp

DISCUSSION

Click on a bubble to navigate the e-poster

introduction
methods
results
discussion

Unified Coding

Although a user should rarely type the SNOMED CT IT(SCTID), experience suggests that from time to time this will happen. A user may also copy and paste an SCTID. There is a significant risk of errors in these processes and inclusion of a check-digit is intended to detect the commonest types of error and reduce the risk of such errors passing undetected. The choice of check-digit algorithm has been made to maximize the detection of common typographical errors. SNOMED CT is using Verhoeff check, which is a Dihedral D5 Check.

Dealing with decimal numbers, there a detailed categorization of the sorts of errors humans make based on a study of 12000 errors (see Table 4):

There is a webpage of SNOMED CT confluence, showing mathematical description of Verhoeff Check. Java Script and Visual Basic codes for Check-digit computation were provided.

CRDO ID followed Verhoeff check algorithm for catching all single errors, all adjacent transpositions, over 95% of twin errors, over 94% of jump transpositions and jump twin errors, and most phonetic errors according to SNOMED CT.

Table 4 Errors Categorization

No.	Error types	Example	Possibility
1	single errors	a → b	60% to 95%
2	omitting or adding a digit	abc → ac	10% to 20%
3	adjacent transpositions	ab → ba	10% to 20%
4	twin errors	aa → bb	0.5% to 1.5%
5	jump transpositions	acb → bca	0.5% to 1.5%
6	jump twin errors	aca → bcb	below 1%
7	phonetic errors	a0 → 1a	0.5% to 1.5%

Knowledge Linkage for CDSS

One main scenario of CRDO is the knowledge linkage for rare diseases. CRDO will act as a hub connecting the diagnosis or conditions with the items of knowledge bases. We tried to connect CRDO with RGO, which reuse the existing knowledge within RGO. Some concepts with the same clinical meaning have different names in different terminologies. For example, Multiple Acyl-CoA Dehydrogenase Deficiency in CRDO has the same meaning with Glutaric aciduria, type 2 in RGO, which is impossible for computers to know. It is also hardly for human being to know that if they don't have the domain knowledge. Fortunately, those two terms are the synonyms of Glutaric aciduria, type 2 (disorder). The description of SNOMED CT will help the linkage between the concepts in CRDO and RGO. In RGO, Glutaric aciduria type 2 may cause renal cystic disease.

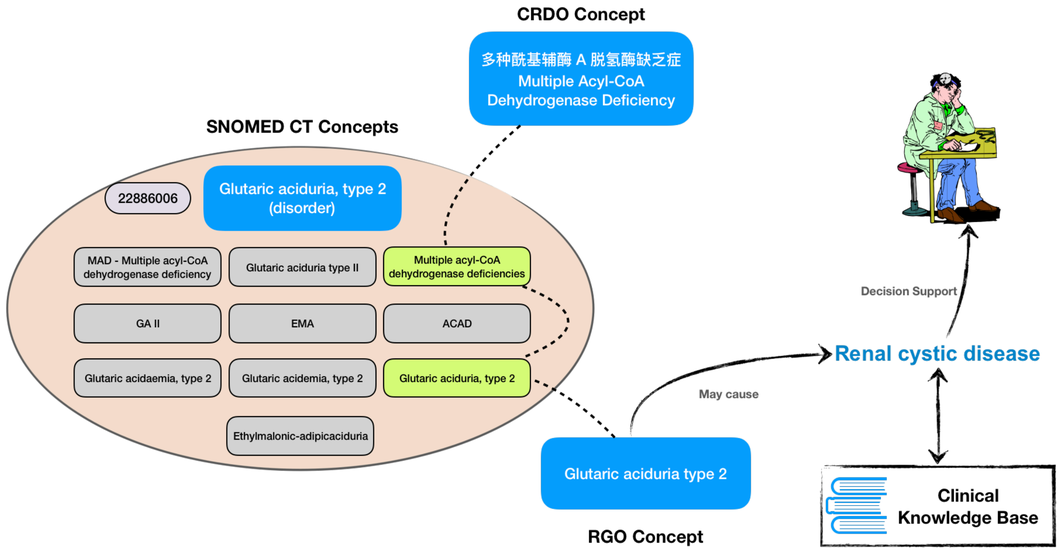

Figure 3 The knowledge linkage by using the mapping between CRDO, SNOMED CT and RGO.