



Using Semantic Web Technology To Support ICD-11 Textual Definitions Authoring

Guoqian Jiang, PhD

Mayo Clinic College of Medicine, Rochester, MN, USA
SWAT4LS 2011, London, UK
December 9, 2011

Acknowledgement

- Co-authors
 - Harold Solbrig
 - Christopher Chute, MD, Dr.PH
- Technical support
 - Cory Endle
- Division of Biomedical Statistic & Informatics,
Mayo Clinic College of Medicine

Type the footnote/source in this space

Background

- The International Classification of Diseases (ICD)
 - An international standard that is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records.

Background

- March 2007
 - The World Health Organization (WHO)
 - Launched the 11th revision of International Classification of Diseases (ICD-11)
- Before May 2012
 - The alpha phase of the ICD-11
 - To augment rubric definitions and review of elements in the Foundation Component of the ICD
- From May 2012
 - The beta phase of the ICD-11
 - Intends to accept public input through a distributed model of crowdsourcing

Type the footnote/source in this space

Textual definitions

- One of core use cases is to create the textual definitions for each ICD category.
- WHO description
 - Each ICD concept will be accompanied by a written definition of its descriptive characteristics. This full text definition enables human users to understand the meaning of a concept for classification, translation and other reasons

Main parameters in ICD-11 content model

- 1 ICD Entity Title
- 2 Classification Properties
- 3 Textual Definitions
- 4 Terms
- 5 Body System/Structure Description
- 6 Temporal Properties
- 7 Severity of Subtype Properties
- 8 Manifestation Properties
- 9 Causal Properties
- 10 Functioning Properties
- 11 Specific Condition Properties
- 12 Treatment Properties
- 13 Diagnostic Criteria

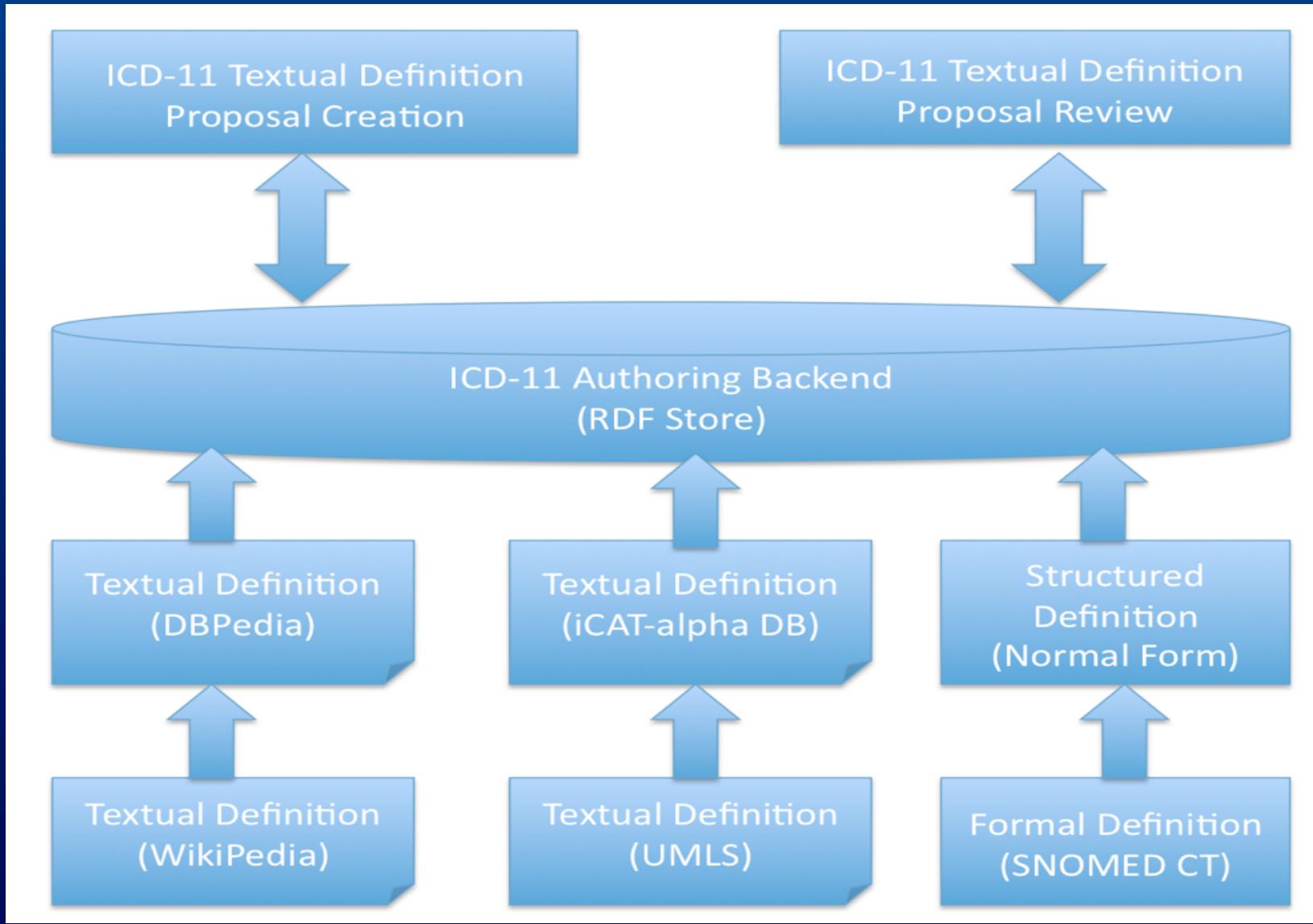
Textual definitions vs Formal definitions

- A well-structured human-readable definition, by distinguishing one entity from another, may serve as the basis for the formal definition (i.e. a computational definition of a class or category, usually expressed in description logic) of an entity.
- While human readable definitions may be more complete and detailed than formal definitions, there should still not be any discordance between them.

Objective of the study

- To design, develop and evaluate approaches to support ICD-11 textual definitions authoring using Semantic Web technology.
- We investigated a number of heterogeneous resources related to the definitions of diseases, including
 - the linked open data (LOD) from DBpedia,
 - the textual definitions from the UMLS and
 - the formal definitions of SNOMED CT.

Proposed system design



Textual definitions from DBpedia

- <http://dbpedia.org/sparql>
- MeSH id as an anchor to map with those corresponding codes in both the SNOMED CT and the ICD-10 through the UMLS CUIs.

```
SELECT DISTINCT ?label ?abstract ?meshId ?wikipediaLink
WHERE {
    ?s a <http://dbpedia.org/ontology/Disease> .
    ?s rdfs:label ?label .
    ?s <http://dbpedia.org/ontology/abstract> ?abstract .
    ?s <http://dbpedia.org/ontology/meshId> ?meshId .
    ?wikipediaLink <http://xmlns.com/foaf/0.1/primaryTopic> ?s .
    FILTER (langMatches(lang(?label), "en") && langMatches(lang(?abstract), "en"))
}
```

Textual definitions from the UMLS

- Imported into the original ICD-11 alpha database for a portion of ICD categories
- Example: ICD category A19 – Miliary tuberculosis

```
SELECT DISTINCT ?label ?definitionContent ?ontologyId ?termId
{ GRAPH <http://who.int/icd>
{ <http://who.int/icd#A19> <http://who.int/icd#definitionPrefilled> ?prefilledDefinition .
<http://who.int/icd#A19> rdfs:label ?label .
?prefilledDefinition <http://who.int/icd#label> ?definitionContent;
<http://who.int/icd#ontologyId> ?ontologyId;
<http://who.int/icd#termId> ?termId .
}}
```

Textual definitions from the UMLS

- Imported into the original ICD-11 alpha database for a portion of ICD categories
- Example: ICD category A19 – Miliary tuberculosis

label	definitionContent	ontologyId	termId
A19. Miliary tuberculosis	An acute form of TUBERCULOSIS in which minute tubercles are formed in a number of organs of the body due to dissemination of the bacilli through the blood stream.	UMLS/MSH2008_2008_02_04	C0041321

Canonical/Normal form of SNOMED CT

- A normal form is a view that can be generated by maximally decomposing any valid expression through applying a set of logical transformation rules.

Structured definitions from SNOMED CT

- We utilized the data files and the canonical table file of the 20100731 International Release of SNOMED CT.
- We defined simple grammatical rules that render the short canonical form into the structured definition (that is more human-readable)

Acute myocardial infarction
is a Disease
that has Clinical course of Sudden onset AND/OR short duration
that has Associated morphology of Acute infarct
and has Finding site of Myocardium structure

Structured definitions from SNOMED CT

- We mapped the SNOMED CT codes and their corresponding structured definitions with the ICD categories represented by the ICD-10 codes through using the UMLS CUIs.
- We then rendered the mappings and definitions into the RDF triples and loaded them into the ICD-11 RDF store.

```
<http://who.int/icd#I21> <http://who.int/icd#icdCode> "I21";
                           <http://who.int/icd#definitionPrefilled> _:b0672 .
_:b0672 <http://who.int/icd#label>
          "Acute myocardial infarction Is a Disease and
           has Clinical course of Sudden onset AND/OR short duration
           that has Associated morphology of Acute infarct
           and has Finding site of Myocardium structure" ;
<http://who.int/icd#ontologyId> "SNOMED CT" ;
<http://who.int/icd#termId> "C0155626" ;
<http://who.int/icd#sctId> "57054005" .
```

System evaluation

- First, we evaluated the coverage of each definition resource.
- Second, we performed a case study on two example ICD categories.
 - We profiled the definitions of each example using the ICD-11 content model.
 - The purpose of this evaluation is to illustrate the potential gap between the textual definitions and the formal definitions.

Results

- From DBpedia
 - 2735 distinct disease definitions and labels
 - 2463 distinct MeSH Ids
 - 1069 ICD categories in ICD-10 codes
- From UMLS
 - 1487 textual definitions for 1278 distinct ICD categories.
 - From 7 different coding schemes of the UMLS
 - NCI Thesaurus, MeSH, and Gene Ontology, etc.

Results

- From SNOMED CT
 - 96235 SNOMED CT concept ids from the branch of “Clinical Finding”
 - Utilizing the grammatical rules we defined, we were able to transform the short canonical forms into the structured definition for each of the codes
 - We mapped 5778 ICD categories represented by the ICD-10 codes to 6122 SNOMED CT concept ids

ICD Categories

- + A39.5. Meningococcal heart disease
- + B37.6. Candidal endocarditis
- + B57.0. Acute Chagas' disease with heart involvement
- + I30. Acute pericarditis
- + I31. Other diseases of pericardium
- + I32. Pericarditis in diseases classified elsewhere
- + I33. Acute and subacute endocarditis
- + I34. Nonrheumatic mitral valve disorder
- I35. Nonrheumatic aortic valve disorder
 - + I35.0. Aortic (valve) stenosis
 - + I35.1. Aortic (valve) insufficiency
 - + I35.2. Aortic (valve) stenosis with insufficiency
 - + I35.8. Other aortic valve disorders
 - + I35.9. Aortic valve disorder, unspecified
- + I36. Nonrheumatic tricuspid valve disorders
- + I37. Pulmonary valve disorders
- + I38. Endocarditis, valve unspecified
- + I39. Endocarditis and heart valve disorders
- + I40. Acute myocarditis

Content Browser

FSN	Definition	External References	Proposal Changes
I35.1. Aortic (valve) insufficiency	backflow of blood from the aorta into the left ventricle, owing to imperfect functioning of the aortic semilunar valve.	UMLS/CSP2006	
I35.1. Aortic (valve) insufficiency	Pathological condition characterized by the backflow of blood from the ASCENDING AORTA back into the LEFT VENTRICLE, leading to regurgitation. It is caused by diseases of the AORTIC VALVE or its surrounding tissue (aortic root).	UMLS/MSH2008	
I35.1. Aortic (valve) insufficiency	The backward flow of blood from the aorta into the left ventricle, owing to insufficiency of the aortic semilunar valve; it may be chronic or acute.	UMLS/NCI2007_0	
I35.1. Aortic (valve) insufficiency	Dysfunction of the aortic valve, with incomplete valve closure resulting in valvular regurgitation.	UMLS/NCI2007_0	
Aortic insufficiency	Aortic insufficiency (AI), also known as aortic regurgitation (AR), is the leaking of the aortic valve of the heart that causes blood to flow backward (regurgitate) during ventricular diastole, from the aorta into the left ventricle. This occurs due to abnormalities of either the aortic valve itself or the surrounding structures of the aorta.	Propose Changes... Create New Definition...	DBpedia
I35.1	Aortic valve regurgitation is a Heart valve regurgitation and Finding site Aortic valve structure		SNOMED CT

Instructions

Proposal Creation

I35.0 Aortic (valve) stenosis

Source	Definition type	Definition	ICD-11 parameters involved
DBpedia	Textual definition	Aortic valve stenosis (AS) is a type of valvular heart disease characterized by an abnormal narrowing of the aortic valve opening.	Supertype, Morphology, anatomical site
UMLS/MS H2008_2008_02_04	Textual definition	A pathological constriction that can occur above (supravalvular stenosis), below (subvalvular stenosis), or at the AORTIC VALVE. It is characterized by restricted outflow from the LEFT VENTRICLE into the AORTA.	Supertype, Morphology, Anatomical site
UMLS/NCI 2007_05E	Textual definition	Narrowing of the orifice of the aortic valve or of the supravalvular or subvalvular regions.	Morphology, Anatomical site
UMLS/CSP 2006	Textual definition	constriction in the opening of the aortic valve or of the supravalvular or subvalvular regions.	Morphology, Anatomical site
SNOMED CT	Structured definition	Aortic valve stenosis is a Disease that has Associated morphology of Stenosis and has Finding site of Aortic valve structure.	Supertype, Morphology, Anatomical site

N17-N19. Renal failure

Source	Definition type	Definition	ICD-11 parameters involved
DBpedia	textual definition	<p>Renal failure or kidney failure (formerly called renal insufficiency) describes a medical condition in which the kidneys fail to adequately filter toxins and waste products from the blood. The two forms are acute and chronic; a number of other diseases or health problems may cause either form of renal failure to occur.</p> <p>Renal failure is described as a decrease in the glomerular filtration rate.</p> <p>Biochemically, renal failure is typically detected by an elevated serum creatinine level. Problems frequently encountered in kidney malfunction include abnormal fluid levels in the body, deranged acid levels, abnormal levels of potassium, calcium, phosphate, and (in the longer term) anemia. Depending on the cause, hematuria (blood loss in the urine) and proteinuria (protein loss in the urine) may occur. Long-term kidney problems have significant repercussions on other diseases, such as cardiovascular disease.</p>	Supertype, Anatomical site, Causal, Manifestation, Diagnostic criteria
UMLS/ MSH20 08_2008 _02_04	textual definition	A severe irreversible decline in the ability of kidneys to remove wastes, concentrate URINE, and maintain ELECTROLYTE BALANCE; BLOOD PRESSURE; and CALCIUM metabolism. Renal failure, either acute (KIDNEY FAILURE, ACUTE) or chronic (KIDNEY FAILURE, CHRONIC), requires HEMODIALYSIS.	Supertype, Anatomical site, Severity, Temporal, Treatment
UMLS/ NCI200 7_05E	textual definition	Acute or chronic condition, characterized by the inability of the kidneys to adequately filter the blood substances, resulting in uremia and electrolyte imbalances. Acute renal failure is usually associated with oliguria or anuria, hyperkalemia, and pulmonary edema. Chronic renal failure is irreversible and requires hemodialysis.--2004	Supertype, Anatomical site, Manifestation, Temporal, Treatment
UMLS/ CSP200 6	textual definition	inability of a kidney to excrete metabolites at normal plasma levels under conditions of normal loading or inability to retain electrolytes under conditions of normal intake.	Anatomical site, Diagnostic criteria
SNOME D CT	structured definition	Renal failure syndrome is a Renal impairment.	Supertype

Case study results

- The textual definitions were more detailed than the structured definitions derived from the formal definitions.
- We found that most of definitions specified the supertypes but the supertypes varied in different granularity.
 - Taking the example from Table 1, the supertypes specified for the “Aortic valve stenosis” include “a valvular heart disease”, “a pathological constriction” or “a disease”.

Discussions

- The textual definitions extracted from DBpedia are a typical example of traditional human readable definitions that are generated using a crowdsourcing model.
- We found that the type “Disease” and the predicate “meshId” defined in DBpedia ontology are very useful for the extraction process.
- The multilingual definitions are available for most of disease instances in DBpedia.
- ICD-11 project may potentially take advantage of the crowdsourcing model of Wikipedia.

Definitions from UMLS

- An ICD code can have multiple definitions from multiple coding schemes identified.
- The definitions may have been authored in different contexts for the different purpose.
- We consider that this is an important source of well-curated definitions that can be used as the references for the ICD-11 use case.

Definitions from SNOMED CT

- We developed an approach to generate the structured definitions from the formal definitions of SNOMED CT concept codes
- We consider that the structured definition derived from the canonical form would represent well the core meaning of the corresponding concept code

In Summary

- Semantic Web technology provides a scalable framework to allow the successful integration of the heterogeneous definition resources in support of ICD-11 textual definitions authoring.

Future work

- Evaluation of the crowdsourcing model of Wikipedia/DBpedia for ICD-11 textual definitions from public input;
- Development of a user interface that allows the ICD community to leverage the definition resources to produce the solid textual definitions for the ICD-11 categories;
- Development of methods and tools to allow the system to check the consistency between the ICD-11 textual definitions and their formal definitions;

Questions?

