Observational Health Data Sciences and Informatics (OHDSI)

Natural Language Processing (NLP) Working Group
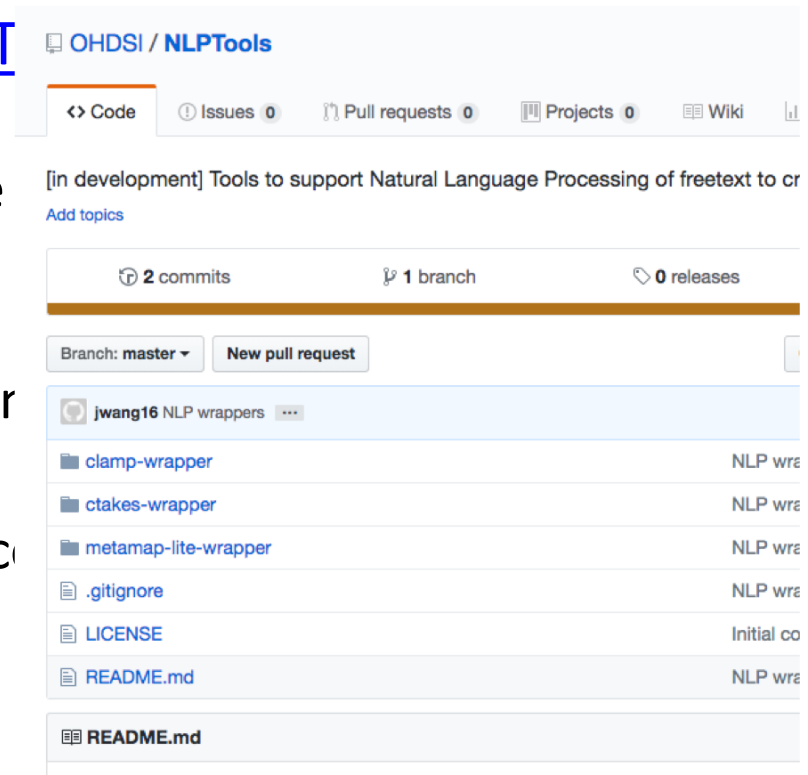
Hua Xu
2/14/2018

AGENDA

- ❖ Introduction of new members
- ❖ GitHub repository information
- ❖ Updates on ongoing projects
  - ❖ Mapping of Note Types to LOINC/standard vocabulary –Karthik Natarajan, Ruth Reeves, Jon Duke and Hua Xu– Report type list discussion
  - ❖ Landscape Analysis of section identifier systems and proposal of a standard terminology for use – Hua Xu, Karthik Natarajan
  - ❖ Mapping of CUIs to standard terminology – Juan Banda, Juan will present some updates today
  - ❖ Rules for defining term_exists – led by Stephane Meystre– Ongoing work, will present in the next meeting
  - ❖ Examples and rules for term_temporal – led by George Hripsack (Sunny) – Sunny to update regarding temporal modifiers.
  - ❖ Standardization of term_modifiers and values – Hua Xu

NLP Wrappers - CLAMP / cTAKES / Metamap Lite

- https://github.com/OHDSI/NLPT
  - Clone source code
  - Import existing project to Eclipse
  - Download dependencies
  - Setup UMLS account
  - Apply CLAMP Command line licer
- Ongoing
  - Mappings from CUI to standard c
  - Normalization of modifiers
  - ...

Mapping from UMLS CUI into the OHDSI Vocabulary

Juan M. Banda

What are we trying to do?

Concept Unique Identifiers (CUI)
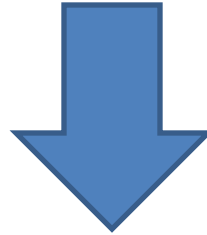
OHDSI concept_id's

ATHENA Standardized Vocabularies

Challenges

- UMLS (2017AB) has 131 English vocabularies

- OHDSI Vocabulary (Jan-2018) has 71+ vocabularies

- Some versioning issues

The Real Challenge

- OHDIS concept_code which is the source code (in RxNorm, SNOMED) can be overlapping between different vocabularies

- Filtering on a per-vocabulary basis is needed

Our approach

1) Identify matching vocabularies on both sides (HPCS, ICD9, ICD10, SNOMED are easy)

UMLS2017AB – MRSAB table

RSAB ▲	SON	SVER	SLC	SSN	SCIT
HCPT	HCPCS Version of Current Procedural Terminology (CPT),...	2017	Cynthia Hake;CMS HCPCS Workgroup Chair;Centers for ...	CPT in HCPCS	;;;;Version of Physicians' Current Procedural Terminology...
HGNC	HUGO Gene Nomenclature Committee, 2017_05	2017_05	Elspeth Bruford, PhD;Group Co-ordinator;HUGO Gene No...	HUGO Gene Nomenclature Committee	Gray KA, Yates B, Seal RL, Wright MW, Bruford EA;;HGN...
HL7V2.5	HL7 Vocabulary Version 2.5, 2003_08_30	2003_08_30	Health Level Seven;;;3300 Washtenaw Avenue,;Suite 22...	HL7 Version 2.5	Mark McDougall, Executive Director, Health Level Seven; ...
HL7V3.0	HL7 Vocabulary Version 3.0, 2016_07	2016_07	Health Level Seven International;;;3300 Washtenaw Ave...	HL7 Version 3.0	;;;;Health Level Seven (HL7) Vocabulary;;;;Ann Arbor, M...
HPO	Human Phenotype Ontology, 2017_06_28	2017_06_28	Dr. Peter N. Robinson;Professor of Computational Biolog...	HPO	;;;;The Human Phenotype Ontology (HPO);;;;;The Huma...
ICD10	ICD10, 1998	1998	;;Office of Publications, World Health Organization;1211 ...	ICD-10	;;;;International Statistical Classification of Diseases and ...
ICD10AE	ICD10, American English Equivalents, 1998	1998	;;Office of Publications, World Health Organization;1211 ...	ICD-10 Am Engl	;;;;International Statistical Classification of Diseases and ...
ICD10AM	International Statistical Classification of Diseases and Rel...	2000	;;National Centre for Classification in Health, University o...	ICD-10 Austral Mod	;;National Centre for Classification in Health, University o...
ICD10AMAE	International Statistical Classification of Diseases and Rel...	2000	;;National Centre for Classification in Health University of...	ICD-10 Austral Mod Am Engl	;;;;International Statistical Classification of Diseases and ...
ICD10CM	International Classification of Diseases, 10th Edition, Clini...	2018	Donna Pickett;Medical Classification Administrator;Nation...	ICD-10-CM	;;National Center for Health Statistics (NCHS), under aut...
ICD10PCS	ICD-10-PCS, 2018	2018	Pat Brooks;Senior Technical Advisor;Center for Medicare ...	ICD-10-PCS	;;Centers for Medicare and Medicaid Services;;Internatio...
ICD9CM	International Classification of Diseases, Ninth Revision, C...	2014	Patricia Brooks;Contact for Procedures;Health Care Fina...	ICD-9-CM	;;National Center for Health Statistics (NCHS);;ICD-9-CM...
ICF	International Classification of Functioning, Disability and	2003_12_10	Nenad Kostanjsek;Technical Officer, Classification, Termi...	International Classification of Functioning, Disability and	;;World Health Organization;International Classification

OHDSI Vocabulary – Vocabulary table

HCPCS	Healthcare Common Procedure Coding System (CMS)	2017 Alpha Numeric HCPCS File
HES Specialty	Hospital Episode Statistics Specialty (NHS)	Not implemented yet
ICD10	International Classification of Diseases, Tenth Revision (...	2016 Release
ICD10CM	International Classification of Diseases, Tenth Revision, ...	ICD10CM FY2018 code descriptions
ICD10PCS	ICD-10 Procedure Coding System (CMS)	ICD10PCS 20160518
ICD9CM	International Classification of Diseases, Ninth Revision, C...	ICD9CM v32 master descriptions
ICD9Proc	International Classification of Diseases, Ninth Revision, C...	ICD9CM v32 master descriptions
ICDO3	ICD-O-3	ICDO3 SEER Site/Histology Released 09/18/2015
ISBT	ISBT 128 Product Description Code Database	7.9.0
ISBT Attribute	ISBT 128 Product Description Code Database	7.9.0
LOINC	Logical Observation Identifiers Names and Codes (Regen...	LOINC 2.61
MDC	Major Diagnostic Categories (CMS)	2013-01-06
Meas Type	OMOP Measurement Type	
MedDRA	Medical Dictionary for Regulatory Activities (MSSO)	MedDRA version 19.1
MeSH	Medical Subject Headings (NLM)	2016 Release

Our approach (2)

2) Obtain CUI via concept_code (OHDSI) and code (UMLS)

CUI	concept_id	vocabulary_id
C0000039	1592753	RxNorm
C0000266	19041783	RxNorm
C0000294	1354698	RxNorm
C0000378	44814542	RxNorm
C0000392	19123344	RxNorm
C0000464	19016670	RxNorm
C0000473	19018384	RxNorm

UMLS via CUI

CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI	SAB	TTY	CODE	STR	SRL	SUPPRESS	CVF
C0000039	ENG	P	L0000039	PF	S17175117	N	A28315139	9194921	1926948	(NULL)	RXNORM	IN	1926948	1,2-dipalmitoylphosphatidylcholine	0	N	(NULL)

OHDSI Vocab via concept_id

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
1592753	1,2-dipalmitoylphosphatidylcholine	Drug	RxNorm	Ingredient	S	1926948	2017-08-07	2099-12-31	

Sample queries

```
#### RXNORM mappings #####
SELECT AA.* FROM (
SELECT A.CUI, B.concept_id, B.vocabulary_id FROM
umls2017ab_AC.MRCONSO as A LEFT JOIN OHDSIVocab.concept
as B ON A.CODE=B.concept_code WHERE A.LAT='ENG' AND
A.SAB='RXNORM' AND B.vocabulary_id='RxNorm'
) AA GROUP BY AA.CUI;


#### CPT4 mappings #####
SELECT AA.* FROM (
SELECT A.CUI, B.concept_id, B.vocabulary_id FROM
umls2017ab_AC.MRCONSO as A LEFT JOIN OHDSIVocab.concept
as B ON A.CODE=B.concept_code WHERE A.LAT='ENG' AND
A.SAB='CPT' AND B.vocabulary_id='CPT4'
) AA GROUP BY AA.CUI;
```

What is next?

- We will share a fully mapped table

- We will then build a mapping package, which has the flexibility of changing the Source-Destination vocabularies (for vocabulary/UMLS refreshes

Question?

Thank you for your time!

jmbanda@stanford.edu

SUTime

- Part of the Stanford CoreNLP pipeline, SUTime is a library for recognizing and normalizing time expressions
- Rule-based system using patterns
- Annotation follows TimeML TIMEX3 standard
- Main temporal types
  - Time instance (e.g. 2011 Aug 11)
  - Duration (e.g. 3 days),
  - Range, a time interval with start and end points
  - Set of temporal, e.g. Every Friday

  - Can also parse relative times with respect to reference date

Ref: https://nlp.stanford.edu/software/sutime.html
https://github.com/evandrix/stanford-corenlp/blob/master/sutime/defs.sutime.txt

SUTime

- Date
  - <TIMEX3 tid="t1" type="DATE" value="1963-10" >October of 1963</TIMEX3>
  - <TIMEX3 tid="t1" type="DATE" value="1994-WI" >winter of nineteen ninety four</TIMEX3>
  - <TIMEX3 tid="t2" type="DATE" value="PAST_REF">recently</TIMEX3>
  - <TIMEX3 tid="t5" type="DATE" value="PRESENT_REF">now</TIMEX3>
  - <TIMEX3 tid="t10" type="DATE" value="FUTURE_REF">future</TIMEX3>
- Duration
  - <TIMEX3 tid="t1" TYPE="DURATION" VAL="P56Y">fifty six years</TIMEX3>
- Duration Range
  - <TIMEX3 tid="t1" type="DURATION" alt_value="P2M/P3M" >two to three months</TIMEX3>
- Set
  - <TIMEX3 tid="t1" value="XXXX-WXX-7" type="SET" quant="every third" periodicity="P3W">Every third Sunday</TIMEX3>
- Time
  - <TIMEX3 tid="t1" value="2011-08-01T17:05:00" type="TIME">5:05 in the afternoon</TIMEX3>