

# ETL for mapping textual data to OMOP NLP table

OHDSI NLP WG Monthly Meeting

12/13/2017

# NOTE & NOTE\_NLP Table

- We have done: wrappers for parsing general clinical NLP systems' (MetaMap, cTAKES, CLAMP) output to defined fields

Field	Required	Type
note_id	Yes	integer
person_id	Yes	integer
note_date	Yes	date
note_datetime	No	datetime
note_type_concept_id	Yes	integer
note_class_concept_id	Yes	integer
note_title	No	varchar(250)
note_text	Yes	RBDMS dependent text
encoding_concept_id	Yes	integer
language_concept_id	Yes	integer
provider_id	No	integer
note_source_value	No	varchar(50)
visit_occurrence_id	No	integer

Field	Required	Type
note_nlp_id	Yes	Big Integer
note_id	Yes	integer
section_concept_id	No	integer
snippet	No	varchar(250)
offset	No	varchar(50)
lexical_variant	Yes	varchar(250)
note_nlp_concept_id	No	integer
note_nlp_source_concept_id	no	integer
nlp_system	No	varchar(250)
nlp_date	Yes	date
nlp_date_time	No	datetime
term_exists	No	varchar(1)
term_temporal	No	varchar(50)
term_modifiers	No	varchar(2000)

# note\_class\_concept\_id

- A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the HL7 LOINC Document Type Vocabulary classification of the note.
- CDO annotates each document with 5 dimensions
  - **Kind of Document:** Characterizes the general structure of the document at a macro level (e.g. Anesthesia Consent)
  - **Type of Service:** Characterizes the kind of service or activity (e.g. evaluations, consultations, and summaries). The notion of time sequence, e.g., at the beginning (admission) at the end (discharge) is subsumed in this axis. Example: Discharge Teaching.
  - **Setting:** Setting is an extension of CMS's definitions (e.g. Inpatient, Outpatient)
  - **Subject Matter Domain (SMD):** Characterizes the subject matter domain of a note (e.g. Anesthesiology)
  - **Role:** Characterizes the training or professional level of the author of the document, but does not break down to specialty or subspecialty (e.g. Physician)
- ~3000 concepts with various granularity

# Mapping UTP Note Types to LONIC

- 3.7 million outpatients
- 11.9 million notes
- 318 note types
  - 61 type has more than 10,000 notes
  - 183 type has more than 1,000 notes
  - Generally one note type for
    - Broad purpose (notes/progress/directives...)
    - Multiple subject domain (specialty)
    - Multiple Settings (outpatient/emergency/inpatient...)
    - → Note content!

# UTP Notes – Notes Table

	id	de_id	d_documentid	providerid	providename	specialty	d_documenttype	d_physicalmanifestation
1						Pediatrics	Questionnaire/Consents-	TIF
2						Family Medicine	Referrals-	TIF
3						Orthopedic Surgery	Clinical Encounter-	TIF
4						Unassigned	Referrals-	TIF
5						Gastroenterology	Drivers License-	TIF
6						Family Medicine	Visit Summary	NOTEFORM
7						Gastroenterology	Patient Result Letter	NOTEFORM
8						Dermatology	Insurance-	TIF
9						Unassigned	Referrals-	TIF
10						Unassigned	Hospital Notes-	TIF
11						Cardiothoracic and Vascular Surgery	Drivers License-	TIF
12						Internal Medicine	Established	NOTEFORM
13						Family Medicine	Result Note	NOTEFORM
14						Orthopedic Surgery	Registration.-	TIF
15						Unassigned	Orders-	TIF
16						Cardiac Surgery	Drivers License-	TIF
17						Orthopedic Surgery	Prescriptions-	TIF
18						Unassigned	Immunization-	TIF
19						Family Medicine	Insurance-	TIF

# UTP Notes – Notes Table

	id	de_id	d_documentid	d_authorid	d_authomame	d_ownerid	d
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

Staff

- staff\_id
- staff\_role
- ...

# UTP Notes – Notes table

	id	de_id	d_documentid	e_encountertype	e_billinglocation	csi_chartsectionname	csi_chartssubsectionname
1				Appointment	UTP HPB CLINIC	Administrative	
2				Appointment		Administrative	
3				Appointment		Clinic Notes	
4				Image Encounter		Administrative	
5				Appointment	UTP BAYSHORE MULTI-SPECIALT	Administrative	
6				Appointment		Clinic Notes	
7				Result Review		Communication	
8				Appointment		Administrative	
9				Image Encounter		Administrative	
10				Image Encounter		Hospital	
11				Appointment	UTP MHH SE MED PLAZA 1	Administrative	
12				Appointment	UTP HPB CLINIC	Clinic Notes	
13				Result Review		Communication	
14				Appointment	UTP SPORTS MEDICINE INSTITU	Administrative	
15				Image Encounter		Orders	
16				Appointment	UTP MHH MEDICAL PLAZA	Administrative	
17				Rx Renewal		Orders	
18				Appointment		Administrative	
19				Appointment		Administrative	

Count	d_documenttype	Record ArtifactSubject Matter Domain	Role	Setting	Type of Service	Kind		
2213131	CED - Automatic Export	68608-9 Summary Note	Multiple-Clinical	Automated-Multiple	Outpatient, Emergency	Summary	Report	Clinical Exchange Document
1239174	Chart Note	LP261106-1 Summary of Encounters note	Multiple-Clinical	Provider	Outpatient	History and Physical	Note	Brief, control note
1205484	Established	LP57379-7 Admission history and physical note	Multiple-Clinical	Provider	Outpatient	History and Physical	Note	Through, encounter note
980668	Clinical Summary-RTF	68608-9 Summary Note	Multiple-Clinical	Provider	Any (Inpatient, Outpatient, Emergency, ICU...)	Multipurpose - Consultation	Note	A status document for patients current status
668858	Clinical Staff Note	LP175705-5 Action Plan / LP74455-4 Advance directives	Multiple-Clinical	Provider	Outpatient		Action Plan(?)	Task assignment / To Do
604929	Telephone Note	LP260937-0 Telephone Encounter	Multiple-Clinical	Nurse/Physician	Outpatient	Progress	Note	
465363	Patient Result Letter	80563-0 Report (?)	Laboratory (?)	Technician	Outpatient	Summary	Letter	Laboratory results report
463325	Result Note	80563-0 Report (?)	Laboratory (?) / Microbiology	Provider/?	Outpatient/Emergency/Inpatient		Note	Clinical outcome/discharge notes
425462	New Patient	LP57352-4 Initial evaluation note	Multiple-Clinical	Nurse/Physician	Outpatient/Emergency	History and Physical	Note	
390419	Visit Summary	84266-6 Nurse Outpatient Initial evaluation note	Multiple-Clinical	Nurse	Outpatient	History and Physical	Note	
322903	Clinical Note	LP57379-7 Admission history and physical note	Multiple-Clinical	Physician	Outpatient	History and Physical	Note	
196344	Clinical Note - RBJ	78509-7 Orthopaedic surgery Admission history and physical note	Orthopedics	Physician	Outpatient	History and Physical	Note	Richmont Bone and Joint
194311	Nurse Note	34113-1 Nursing facility note	Multiple-Clinical	Nurse	Outpatient/Emergency	Multipurpose-progress/directives		Mixed, mainly with directives
162939	Est Patient	LP57379-7 Admission history and physical note	Multiple-Clinical	Provider	Outpatient	History and Physical	Note	Through, encounter note
134555	Nurse Triage	84276-5 Nursing Urgent Care center triage note	Emergency	Nurse	Emergency	History and Physical	Note	
124157	Pre-Live Dictation.							Various document types, letter, encounter for dictation
121959	Ultrasound Report	68604-8 Radiology Diagnostic Study note	Radiology	Provider		Diagnostic Study Report		
112417	Result Letter	80563-0 Report (?)	Laboratory (?)	Technician	Outpatient	Summary	Letter	Laboratory results report



# Discussion

- Intension is important (within the document – a directive, a progress, or just an information)
  - Outcome summary
    - Letter to referring physician
    - Summary for consulting physician
    - Discharge summary
    - Phone encounter
- New vs. Established differentiated

# section\_concept\_id

- A foreign key to the predefined Concept in the Standardized Vocabularies representing the section of the extracted term.
- **Current status:** Section headers are extracted from cTAKES and CLAMP outputs
- **Remaining work:** Map section headers to standard concept in CONCEPT table

	TERM CONNECTIONS (14)			
Type Concept	RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
Record Artifact	Is a	<a href="#">Record document part</a>	4309090	SNOMED
SNOMED <span>?</span>	Subsumes	<a href="#">Administration section</a>	44804346	SNOMED
4307844		<a href="#">Assessment section</a>	4314936	SNOMED
422813005		<a href="#">Chief complaint section</a>	4307849	SNOMED
Valid		<a href="#">Correspondence section</a>	40482006	SNOMED
Non-standard		<a href="#">Family history section</a>	4307151	SNOMED
Document section Document section (record artifact)		<a href="#">General document section types</a>	44807904	SNOMED

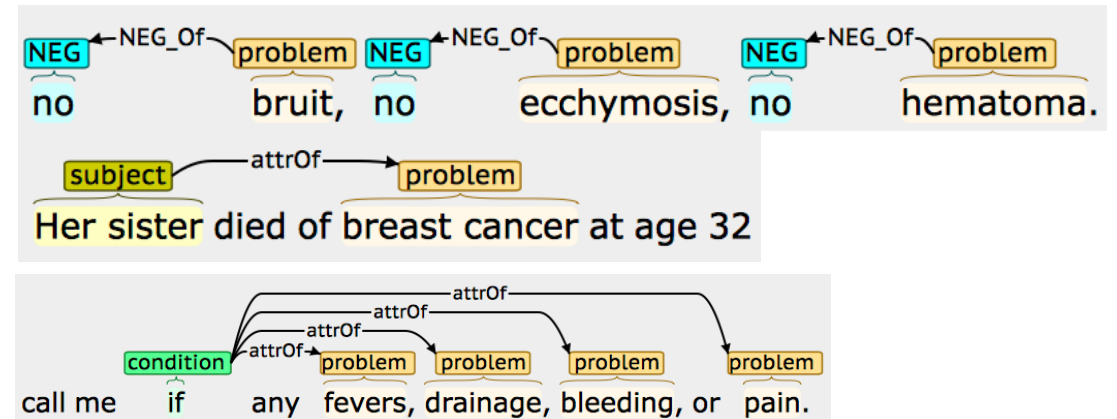
- [10154-3](#) Chief complaint Narrative - Reported
- [10157-6](#) History of family member diseases Narrative
- [10160-0](#) History of medication use Narrative
- [10164-2](#) History of present illness Narrative
- [10183-2](#) Hospital discharge medications Narrative
- [10184-0](#) Hospital discharge physical findings Narrative
- [10187-3](#) Review of systems Narrative - Reported
- [10210-3](#) Physical findings of General status Narrative
- [10216-0](#) Surgical operation note fluids Narrative

# note\_nlp\_concept\_id, note\_nlp\_source\_concept\_id

- 'note\_nlp\_concept\_id', A foreign key to the predefined Concept in the Standardized Vocabularies reflecting the normalized concept for the extracted term. Domain of the term is represented as part of the Concept table.
- 'note\_nlp\_source\_concept\_id', A foreign key to a Concept that refers to the code in the source vocabulary used by the NLP system.
- **Current status:** UMLS CUIs are extracted from cTAKES, MetaMap, CLAMP
- **Remaining work:** convert from UMLS CUIs to standard concept id

# term\_exists

- A summary modifier that signifies presence or absence of the term for a given patient. Useful for quick querying.
- **Current status:**
  - parsed results from 'negation detection' component (cTAKES, MetaMap) and 'ner-attribute' pipeline (CLAMP);
  - term\_exists is **FALSE**, if:
    - 1). Concept is negated;
    - 2). Subject of the concept is not patient;
    - 3). Condition of the concept is not null;
- **Remaining work:** finalize the rules for each system



# term\_temporal

- An optional time modifier associated with the extracted term. (for now “past” or “present” only).
- **Current status:** ‘temporal recognition/relation’ components (CLAMP, cTAKES);
- **Remaining work:** more specific rules determining an event is ‘past’ or ‘present’;

HISTORY: This 67-year old Hispanic male patient was admitted because of enlarged prostate and symptoms of bladder neck obstruction. Physical examination revealed normal

course. He is asked to continue the Colace as an outpatient for stool softening for comfort.

BRIEF HOSPITAL COURSE: The patient is a 71-year-old lady who came in with increased shortness of breath of one day duration.

they recommended further evaluation by an MRI as this could be a hemorrhagic cyst or a solid mass or tumor, so a

# term\_modifiers

- A compact description of all the modifiers of the specific term extracted by the NLP system. (e.g. “son has rash” -> “negated=no,subject=family, certainty=undef, conditional=false, general=false”).
- **Current status:** Parsed modifiers from cTAKES,MetaMap,CLAMP
- **Remaining work:** standardize possible modifiers and their values