



OHDSI NLP WG Updates – 2020.11

by Hua Xu

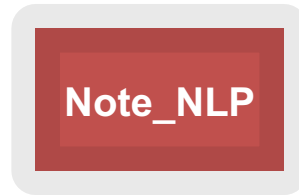
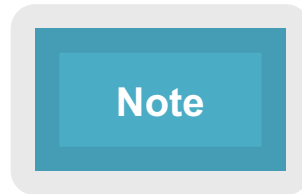
- Recent Activities
 - ETL tool development
 - Note Type Normalization
 - COVID-19 lab test normalization



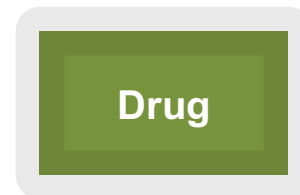
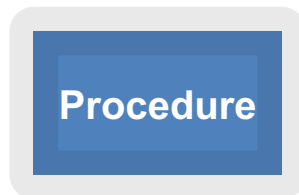
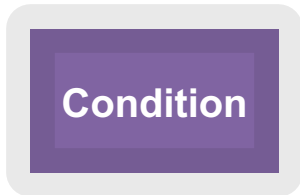
A Potential ETL Workflow for NLP

<https://github.com/OHDSI/NLPTools>

- ! Note Type Normalization
- ! Section Header Normalization
- ! Concept Modifiers



Type_Concept = "NLP derived"



THEIA

- ! Any NLP Engines: cTAKES, MetaMap, CLAMP
- ! Wrappers: Convert NLP outputs to Note_NLP format
- ! Ananke: UMLS to OMOP Concepts
- ! Visualization: search , phenotyping ..



Note Type Normalization (ongoing)

18,075 clinical document titles from five institutions

- Boston Children's Hospital (7,400)
- Vanderbilt University Medical Center (3,434)
- Stanford University Medical School (3,128)
- The University of Texas Health Science Center at Houston (3,232)
- Columbia University Medical Center (881)



LOINC Document Ontology (DO) Axis:

- **Type of Service (ToS):** the kind of healthcare services provided to patients. e.g., Consultation, Evaluation and Management, Procedure
- **Kind of Document (KoD):** the type of clinical documents based on its structure. e.g., Note, Report, Checklist
- **Setting:** the location or channel where clinical care is provided. e.g., Ambulance, Birthing Center, Intensive Care Unit
- **Role:** people and their occupations involved in the service or authors who created the clinical note. e.g., Physicians, Nurse, Pharmacist
- **Subject Matter Domain (SMD):** clinical specialty relevant to the document or the main purpose of creating the document. e.g., Anesthesiology, Urology, Cardiovascular Disease

LOINC DO Axis	Precision		Recall		F1	
	BERT	CRF	BERT	CRF	BERT	CRF
ToS	0.7187	0.7880	0.7848	0.7270	0.7494	0.7120
KoD	0.9076	0.9110	0.9286	0.8930	0.9179	0.9020
Setting	0.8911	0.9190	0.9226	0.8940	0.9058	0.9060
Role	0.8810	0.9210	0.8837	0.8610	0.8811	0.8900
SMD	0.8153	0.8139	0.8434	0.7880	0.8290	0.8000



COVID-19 TestNorm

- An open source tool to map COVID-19 testing names to standard LOINC codes
- Online demo <https://clamp.uth.edu/covid/loinc.php>

COVID-19 Testing Name :

NOVEL CORONAVIRUS 2019 rRT PCR, NASOPHAR

Mapping

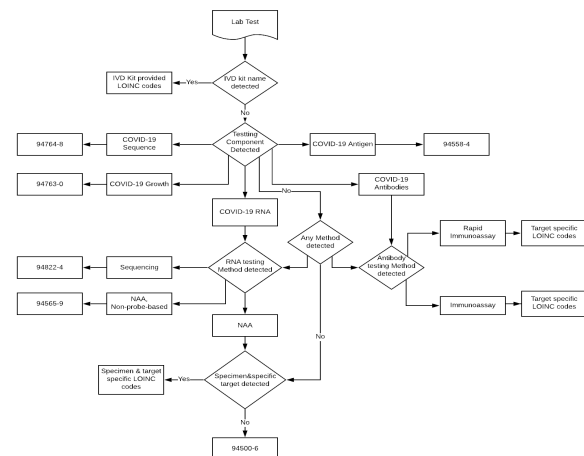
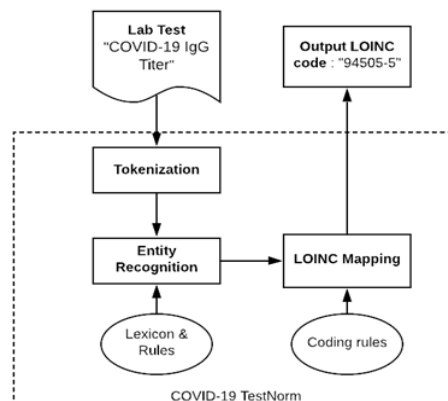
LOINC Code :

94759-8

Long Common Name :

SARS coronavirus 2 (COVID19) RNA [Presence] in Nasopharynx by NAA with probe detection

A fast collaboration among 8 healthcare organizations within a month!





NLP WG Plans

- Update the NLP table (version 2?)
- Continue ETL development
- Conduct cross-site studies that use textual data
 - Descriptive analysis and possible effect estimation of prone positioning in non intubated COVID-19 patients - <https://forums.ohdsi.org/t/new-study-proposal-descriptive-analysis-and-possible-effect-estimation-of-prone-positioning-in-non-intubated-covid-19-patients/12013>
- Open-source python package for clinical NLP (planning)