Recent Activities

- ETL tool development
- Note Type Normalization
- COVID-19 lab test normalization
A Potential ETL Workflow for NLP

Note

Note Type Normalization
Section Header Normalization
Concept Modifiers

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

**Type_Concept** = “NLP derived”

https://github.com/OHDSI/NLPTools
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

<table>
<thead>
<tr>
<th>LOINC DO Axis</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BERT</td>
<td>CRF</td>
<td>BERT</td>
</tr>
<tr>
<td>ToS</td>
<td>0.7187</td>
<td>0.7880</td>
<td>0.7848</td>
</tr>
<tr>
<td>KoD</td>
<td>0.9076</td>
<td>0.9110</td>
<td>0.9286</td>
</tr>
<tr>
<td>Setting</td>
<td>0.8911</td>
<td>0.9190</td>
<td>0.9226</td>
</tr>
<tr>
<td>Role</td>
<td>0.8810</td>
<td>0.9210</td>
<td>0.8837</td>
</tr>
<tr>
<td>SMD</td>
<td><strong>0.8153</strong></td>
<td>0.8139</td>
<td><strong>0.8434</strong></td>
</tr>
</tbody>
</table>
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](https://clamp.uth.edu/covid/loinc.php)

**COVID-19 Testing Name:** NOVEL CORONAVIRUS 2019 rRT PCR, NASOPHAR

**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
- Open-source python package for clinical NLP (planning)