OHDSI NLP Workgroup Updates

July 19, 2022

Hua Xu

OHDSI NLP Workgroup - Objectives

Promote the use of textual information from electronic health records (EHRs) for observational studies under the OHDSI umbrella

1. Develop methods and tools to represent and process clinical text
2. Conduct cross-institutional studies utilizing textual data in EHRs
3. Disseminate best practice of using textual data for real world evidence generation

OHDSI NLP Workgroup – Meeting schedule

Second Wednesday of every month @ *2 PM – 3 PM ET*

Wiki page: https://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:nlp-wg
GitHub repository: https://github.com/OHDSI/NLPTools

Hua.Xu@uth.tmc.edu

OKRs for 2022

1. Objective 1: revise representation of textual data in OMOP CDM (Note_NLP) and develop user guide

2. Objective 2: develop multiple NLP tools to extract, transform, and load textual data into Note_NLP

3. Objective 3: conduct at least two multi-site clinical studies that utilize both structured and textual data

4. Objective 4: continuous community engagement

OKRs 2022 – Objective 1

Objective 1 : revise representation of textual data in OMOP CDM (Note_NLP) and develop user guide

Completed:

- Submitted a paper on representing and utilizing clinical text for observational studies

- Submitted a proposal to the CDM WG for modifications to the Note_NLP table

In the pipeline:

- Schema for intermediate tables to store modifiers, note types and additional NLP features

OKRs 2022 – Objective 2

Objective 2 : develop multiple NLP tools to extract, transform, and load textual data

Key results:

Task	Q1 2022	Q2 2022	Q3 2022	Q4 2022
At least four NLP tools that can directly generate Note-NLP outputs (Hongfang, Tim, Paul, Hua)	Tool 1	Tool 2	Tool 3	Tool 4
Sample scripts for transferring data from Note_NLP table to clinical event tables (Vipina)	Publish on NLP GitHub			
Tool to normalize "note_type" (Karthik)	Define representation	Develop program	Publish on NLP GitHub	Feedback and revise
Release a new version of Ananke (Juan)			Present at symposium	

OKRs 2022 – Objective 3

Objective 3: conduct at least two multi-site clinical studies that utilize both structured and textual data

In progress:

1. Characterizing Post-Acute Sequelae of SARS-CoV-2 (PASC) Infection
   – 4 sites

Objective 4: continuous community engagement

Key results:

– Monthly meetings

– Invited presentations

– Project-specific meetings

– Forum QA on NLP implementation