Identifying Patients for Clinical Trials Using Information Extraction Augmented by Medical Ontologies

Zunaid Kazi, Abdul Hasib Sazzad, Habibullah Araphat, Mir Md Faysal, Shawon Barua, Mutaherul Islam
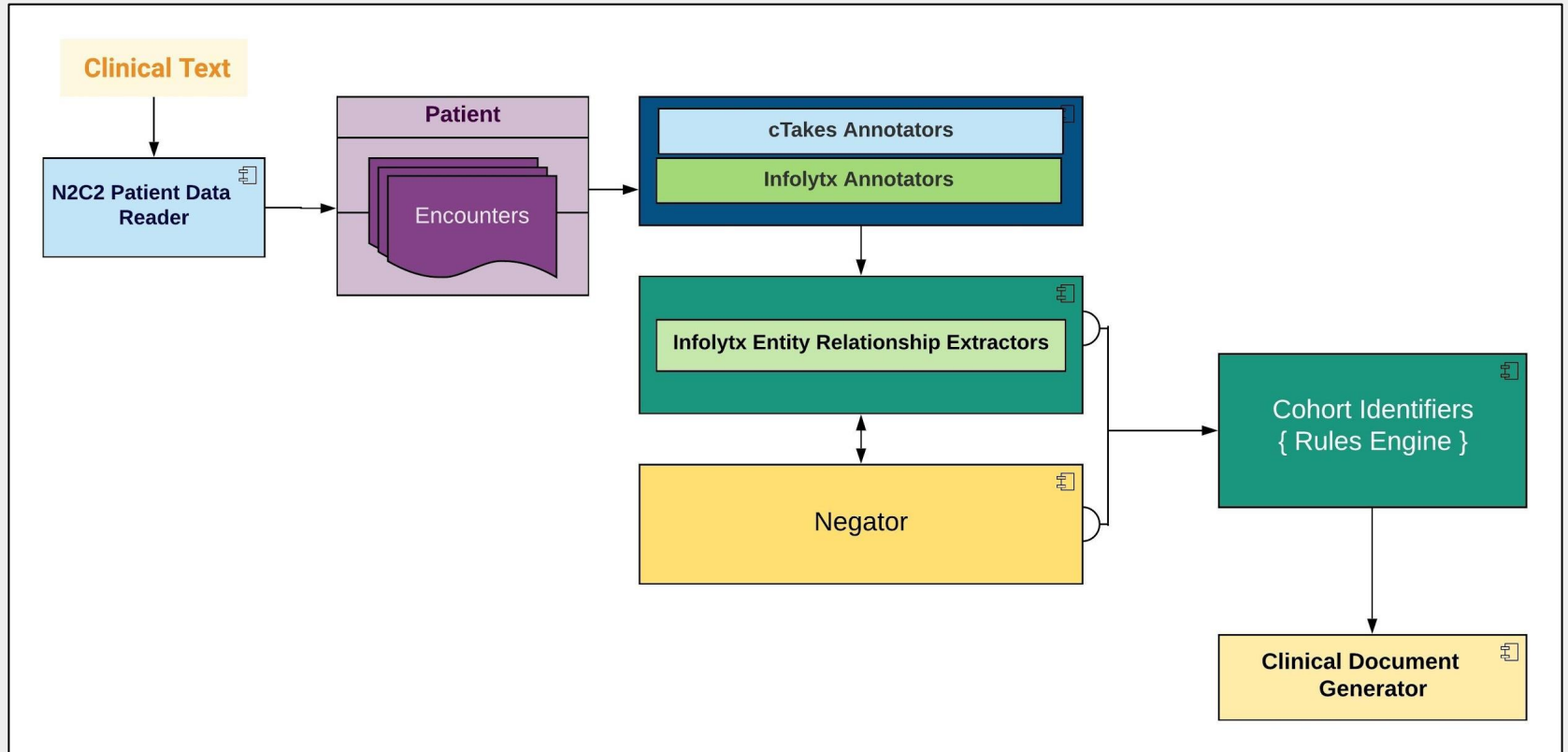
Infolytx

INFOLYTX

Focus

High accuracy annotations of clinical terms from clinical records

High accuracy in inferring and ascribing patient attributes

Architecture

Enhanced Annotators

Simple Segment

Sentence

Tokenizer

Context Dependent Tokenizer

POS Tagger

Chunker

Chunker Adjuster

Language

Creatinine

Drug Abuse

Aspirin

HbA1c

Keto

Abdominal

MI

Why we did as well as we did?

Example: Improved Negation

Multiphrase negation

Sentence: He has no history of cardiac arrest, drug abuse or alcohol.

```
SENTENCE:    He   has no history    of cardiac    arrest ,    drug     abuse     or alcohol .
             PRP  VBZ DT    NN       IN    JJ        NN         NN        NN      CC    NN
                         |======|       |======| |======|    |======| |======|      |======|
                          Finding        Anatomy   Disorder     Drug    Disorder      Drug
```

Example: Improved Negation

Physician English

Sentence: There are no findings to suggest BOWEL OBSTRUCTION.

```
SENTENCE:    There  are  no  findings  to  suggest   BOWEL    OBSTRUCTION.
             EX     VBP  DT  NNS        TO  VB        NN       NN
                            |======|              |======| |=========|
                            Finding               Anatomy   Finding
                            C0037088              C0021853  C0028778
                                                  |=================|
                                                       Disorder
                                                       C0021843
```

Example: Improved Negation

Physician English

Sentence: She denies any drug abuse.

```
SENTENCE:   She denies any    drug    abuse  .
            PRP  VBZ    DT      NN       NN
                 |====|       |======| |======|
                 Event         Drug    Disorder
                            C0013227  C0013146
                                       |======|
                                       Finding
                                       C0562381
```

Example: Integrating ontologies

Dementia MeSH Descriptor Data 2018

Details   Qualifiers   MeSH Tree Structures   Concepts

Nervous System Diseases [C10]
    Central Nervous System Diseases [C10.228]
        Brain Diseases [C10.228.140]
            Acute Febrile Encephalopathy [C10.228.140.021]
            Akinetic Mutism [C10.228.140.042]
            Amblyopia [C10.228.140.055]
            Amnesia, Transient Global [C10.228.140.060]
            Auditory Diseases, Central [C10.228.140.068] ⊕
            Basal Ganglia Diseases [C10.228.140.079] ⊕
            Brain Abscess [C10.228.140.116] ⊕
            Brain Damage, Chronic [C10.228.140.140] ⊕
            Brain Death [C10.228.140.151]
            Brain Diseases, Metabolic [C10.228.140.163] ⊕
            Brain Edema [C10.228.140.187]
            Brain Injuries [C10.228.140.199] ⊕
            Brain Neoplasms [C10.228.140.211] ⊕
            Cerebellar Diseases [C10.228.140.252] ⊕
            Cerebrovascular Disorders [C10.228.140.300] ⊕
            Dementia [C10.228.140.380] ⊖
                AIDS Dementia Complex [C10.228.140.380.070]
                Alzheimer Disease [C10.228.140.380.100]
                Aphasia, Primary Progressive [C10.228.140.380.132] ⊕

Example: Integrating dictionaries

Annotation of Dietary Supplements

Example: Domain Knowledge

He is starting to improve his self-care of dm and last a1c was 8.2

Blood sugars mostly less than 250.

LABORATORY STUDIES: An MRI from September 18 showed a left kidney mass, cystic. Hemoglobin A1c from August was 6.5 and the LDL was 125.

HPI: The pt is a 71 yo female here for follow-up visit. Was unaware of results of May blood draw with increase in HA1C of 8.0.

Metformin 500MG TABLET take 1 Tablet(s) PO BID , Take as directed.

CURRENT MEDICATIONS include Celexa 20 mg daily, Neurontin 600 mg q.i.d., Phenergan p.r.n., Avandia, Glucophage, Lopressor, Zestril, Pravachol, insulin, and aspirin 325 mg daily.

Results

MET

Precision 87.11
Recall 83.01
Fscore 84.55

NOT MET

Precision 88.27
Recall 91.07
Fscore 89.67

| | ----------- met ------------- | | | | ------ not met ------- | | | -- overall --- | |
	Prec.	Rec.	Speci.	F(b=1)	Prec.	Rec.	F(b=1)	F(b=1)	AUC
Abdominal	0.5000	0.7333	0.6071	0.5946	0.8095	0.6071	0.6939	0.6442	0.6702
Advanced-cad	0.8780	0.8000	0.8780	0.8372	0.8000	0.8780	0.8372	0.8372	0.8390
Alcohol-abuse	0.0000	0.0000	0.9639	0.0000	0.9639	0.9639	0.9639	0.4819	0.4819
Asp-for-mi	0.8395	1.0000	0.2778	0.9128	1.0000	0.2778	0.4348	0.6738	0.6389
Creatinine	0.6190	0.5417	0.8710	0.5778	0.8308	0.8710	0.8504	0.7141	0.7063
Dietsupp-2mos	0.8333	0.6818	0.8571	0.7500	0.7200	0.8571	0.7826	0.7663	0.7695
Drug-abuse	0.2500	0.3333	0.9639	0.2857	0.9756	0.9639	0.9697	0.6277	0.6486
English	0.9481	1.0000	0.6923	0.9733	1.0000	0.6923	0.8182	0.8958	0.8462
Hba1c	0.9643	0.7714	0.9804	0.8571	0.8621	0.9804	0.9174	0.8873	0.8759
Keto-1yr	0.0000	0.0000	1.0000	0.0000	1.0000	1.0000	1.0000	0.5000	0.5000
Major-diabetes	0.9630	0.6047	0.9767	0.7429	0.7119	0.9767	0.8235	0.7832	0.7907
Makes-decisions	0.9878	0.9759	0.6667	0.9818	0.5000	0.6667	0.5714	0.7766	0.8213
Mi-6mos	0.4444	0.5000	0.9359	0.4706	0.9481	0.9359	0.9419	0.7063	0.7179
	-------------------------------				-----------------------			--------------	
Overall (micro)	0.8411	0.8301	0.8907	0.8355	0.8827	0.8907	0.8867	0.8611	0.8604
Overall (macro)	0.6329	0.6109	0.8208	0.6141	0.8555	0.8208	0.8158	0.7150	0.7159