Step 1: Convert text files to BioC format

python csv2bioc.py -i ../../example.csv -o ../../example.xml

Step 2: De-identify reports

>>> import nltk

>>> nltk.download('averaged\_perceptron\_tagger')

python deidentify.py -i ../../example.xml -o ../../example\_deid.xml

Step3: split each report into sections

We need to install medspacy first

pip install medspacy

python split\_section.py medspacy -i ../../Pneumonia\_deid.xml -o ../../Pneumonia\_section.xml

step 4: preprocess

we also need to install en\_core\_web\_sm first.

pip install -U pip setuptools wheel

pip install -U spacy

python -m spacy download en\_core\_web\_sm

python preprocess\_pipeline.py spacy -i ../../example\_section.xml -o ../../example\_ud.xml --overwrite

Step 5: Name entity recognition

We also need to install intervaltree.

pip install intervaltree.

We also need to add these two commands into the ner.py file

​​import nltk

nltk.download('stopwords')

need to modify cxr14\_phrases\_v1.yml based on the disease

python ner.py regex --phrases /home/mil4012/RADTEXT/radtext/resources/cxr14\_phrases\_v1.yml -i ../../pneumomediastinum\_ud.xml -o ../../pneumomediastinum\_ner.xml --overwrite

/home/mil4012/RADTEXT/radtext/resources/chexpert\_phrases.yml -i ../../pneumomediastinum\_ud.xml -o ../../pneumomediastinum\_ner.xml --overwrite

step 6 parsing

python parse.py -i ../../example\_ner.xml -o ../../example\_parse.xml

step 7 negation

python neg.py -i ../../example\_parse.xml -o ../../example\_neg.xml --overwrite

step 8

python collect\_neg\_labels.py --phrases /home/mil4012/RADTEXT/radtext/resources/cxr14\_phrases\_v1.yml -i ../../example\_neg.xml -o ../../example\_neg.csv