**Readme for GeoTopic Parser, cTakes Parser and Tesseract OCR**

**GeoTopic Parser**

1. Downloaded [Lucene Geo Gazetteer](http://github.com/chrismattmann/lucene-geo-gazetteer)
2. mvn install assembly:assembly
3. Added allCountries.txt file in Lucene Geo Gazetteer
4. Downloaded and installed NER model
5. Started the tika server to get Geo locations from images and html ads.

java -classpath /Users/maheshkumarlunawat/lucene-geo-gazetteer-master/src/location-ner-model:/Users/maheshkumarlunawat/lucene-geo-gazetteer-master/src/geotopic-mime:tika-server/target/tika-server-1.11.jar org.apache.tika.server.TikaServerCli

1. Using curl command extracted the geolocation metadata:

curl -T /Users/maheshkumarlunawat/lucene-geo-gazetteer-master/src/polar.geot -H "Content-Disposition: attachment; filename=polar.geot" http://localhost:9998/rmeta

**cTakes Parser**

1. Downloaded apache-ctakes-3.2.2-bin.tar.gz and installed
2. Signed up for UMLS account with

Username: pramodnagarajarao

Password: pramodnagarajarao

1. Started the server using following command:

java –classpath /Users/maheshkumarlunawat/ctakes/src/ctakes-config: tika-server/target/tika-server-1.11.jar: /Users/maheshkumarlunawat/ctakes/desc: /Users/maheshkumarlunawat/ctakes/resources: /Users/maheshkumarlunawat/ctakes/lib/\\* org.apache.tika.server.TikaServerCli -- config=$HOME/src/ctakes-config/tika-config.xml

1. Using curl command extracted the cTakes metadata:

curl -T /Users/maheshkumarlunawat/files/1.png -H "Content-Disposition: attachment; filename=1.png" http://localhost:9998/rmeta

**Tesseract OCR**

1. Installed tesseract using “brew install tesseract --all-languages”
2. Started the tika-server using following command:

java -jar /path/to/tika-server-1.11-SNAPSHOT.jar

1. Extracted the text from images using command:

“curl T /Users/maheshkumarlunawat/files/image.tiff http://localhost:9998/tika --header "Content-type: image/tiff"