Gustavo Blanco

Michael De La Torre

CS 172: Intro to Information Retrieval

Prof: Dr. Vagelis Hristidis

TA: Shiwen Cheng

Spring 2014

**\*\*UPDATED VERSION WAS E-MAILED TO YOU**

**Team**

Gustavo Blanco ([gblan002@ucr.edu](mailto:gblan002@ucr.edu))

I worked on setting up Apache Lucene & Tomcat. I developed the web interface using the Foundation templates and the Google Maps API. I also developed the indexer program that takes all of the tweets.

Michael De La Torre ([mdela011@ucr.edu](mailto:mdela011@ucr.edu))

<http://lucene.apache.org/core/>

<http://tomcat.apache.org/>

<https://developers.google.com/maps/documentation/javascript/>

<http://foundation.zurb.com/>

**SYSTEM**

Development Environment:

Java 1.7

Apache Lucene 3.5.0

Apache Tomcat 8.0.8

OS X 10.9.2 & Windows 7

Terminal

Sublime Text Editor 2

 [user@server] sh indexer.sh <input-dir>

Our application breaks down into two programs, our Indexer and out WebApp. The indexer program will simply take in the full path of the directory contacting all of the .JSON files that have the tweets. We are assuming each line is a tweet. The program will then begin to parse through the files in the directory, and then the lines in the files in order to index the individual tweets based on the tweets text, user of tweet, geo coordinates, time stamp, and link title. It will store all the indexed files into a directory names “tweetIndex”; this directory will be in the same directory as the indexer program.

The second part to our application is the WebApp. This WebApp uses a Tomcat server to serve a page with an input field and the map centered at the users geo location based on the browser. Once the users typed something into the input field and hits enter the WebApp will make a call to our searcher program with the query provided by the user. The searcher program will them return a list of relevant tweets and which point the WebApp will parse through that list plotting all the of tweets on the map.

**LIMITATIONS**

One of our programs limitations would be that it is not taking in tweets in real-time. Instead it has only indexed the tweets that we sporadically collected over time and those are the only tweets that will be looked at when retrieving query results.

**INSTRUCTIONS**

**SCREENSHOTS**