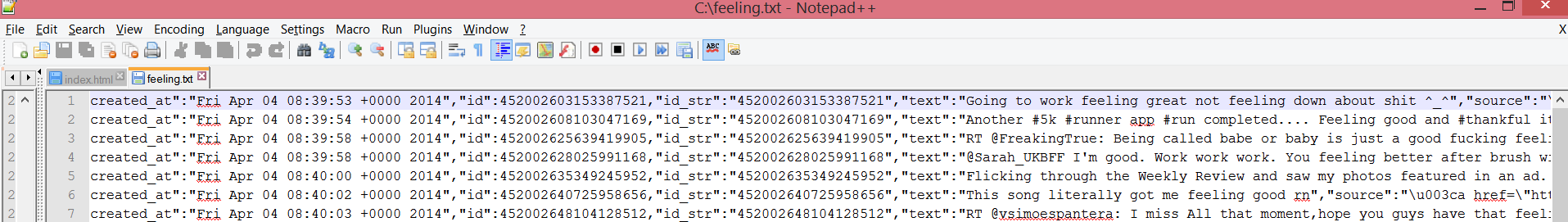
**Title:** World Sentiment Analysis

**1** **Design /Summary:**  A lot of research articles mention that people share their most intimate feeling on social media like twitter and Facebook. So, I decided to analyze what people in different parts of the world are feeling, by analyzing their twitter posts.

**Integration model and algorithm:**

First I captured almost 475 Mb of tweets from all over the world using the Twitter 1.1 public streaming API.

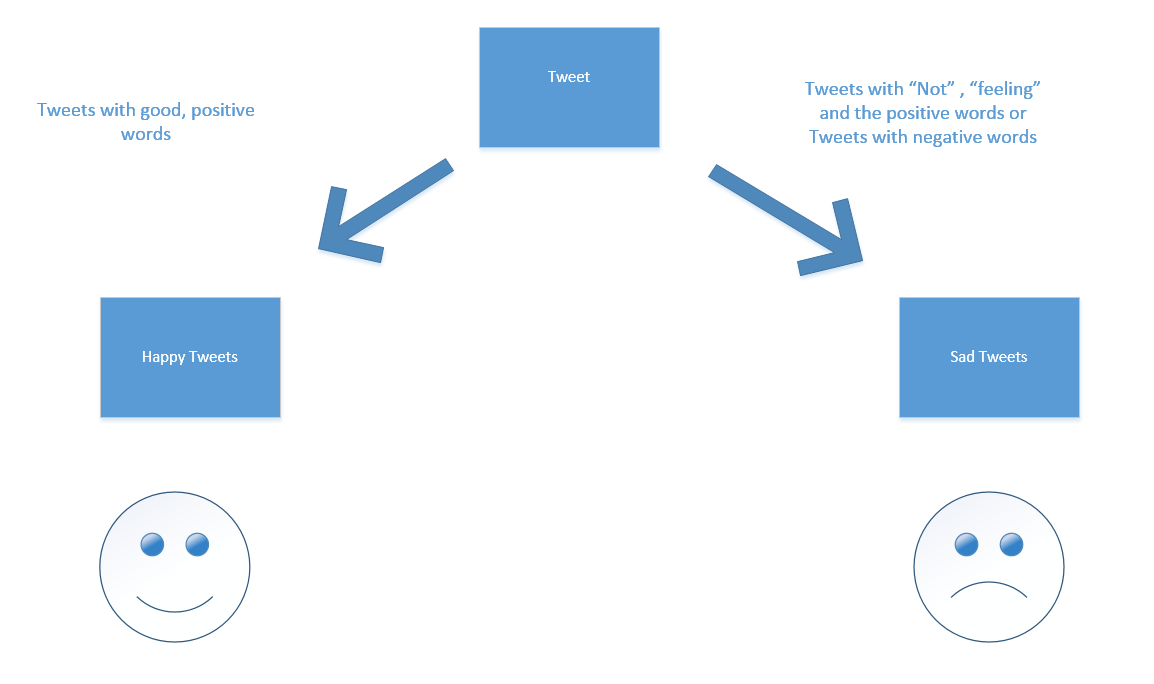


But the tweets contain a lot of redundant information. I then wrote a Map reduce program combined with a classifier, The program takes the 475 Mb text file filled with tweets and then it looks at each and every tweet, If the words in a tweet match a specific set of words or satisfy a specific set of rules then the tweets are categorized as a “Happy Tweet” or a “Sad Tweet”. The program then captures the locations of these tweets and then converts the result into a json format which is ready to push into solar, and the output file size is 15kb.

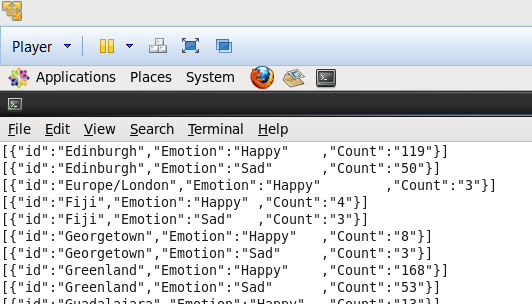
Decision/ Classification Code snippet from the Map Reduce.

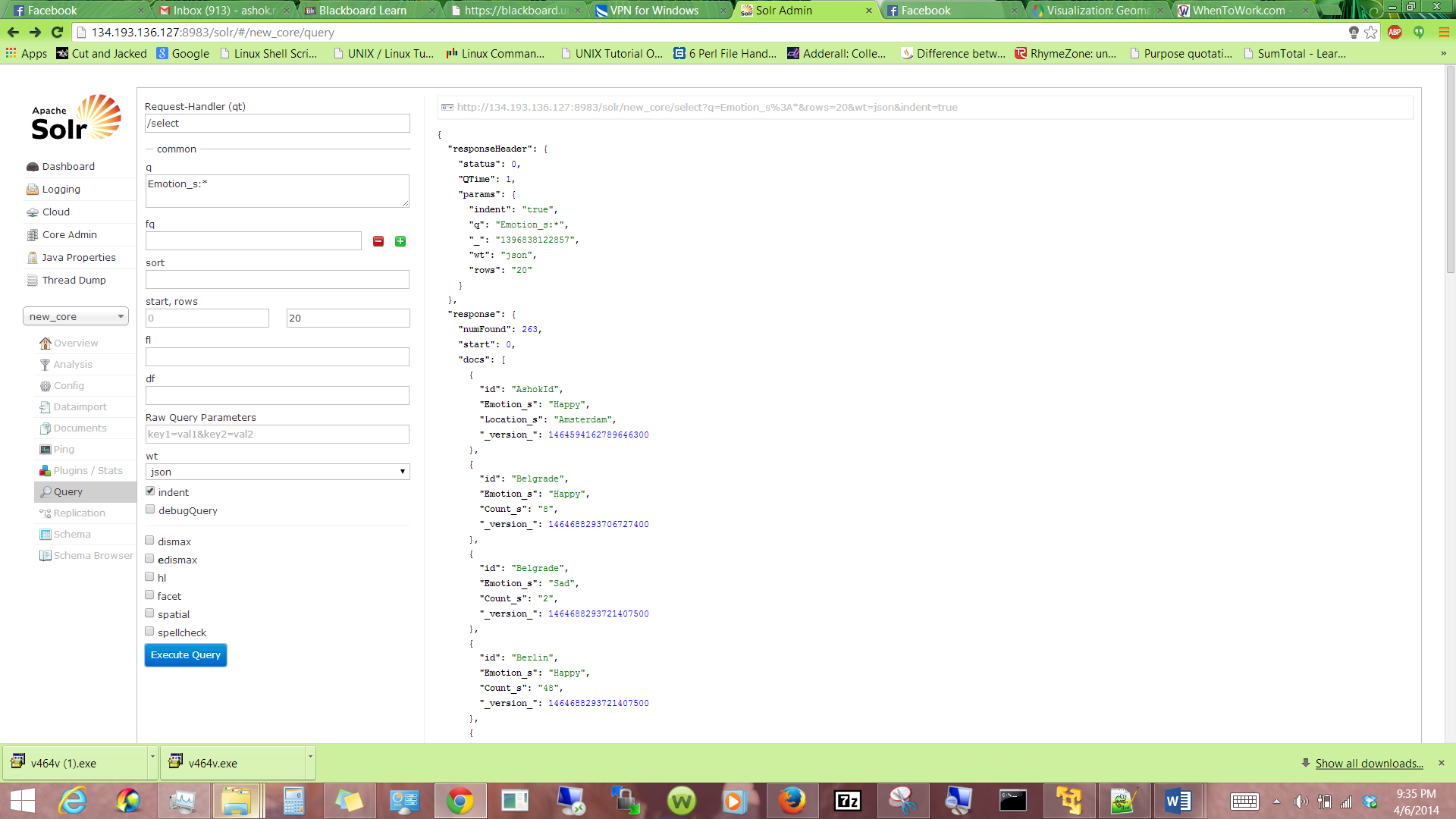


Decision/Classification Algorithm in Map Reduce



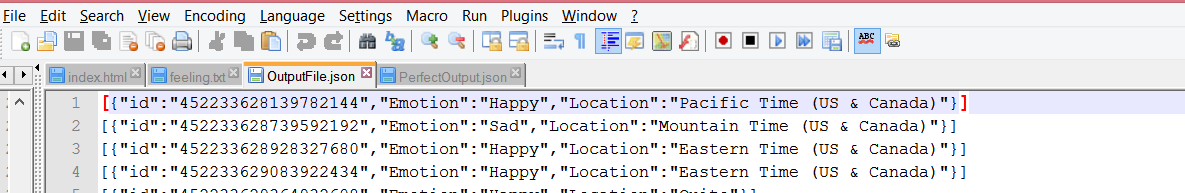
Out put Tweets from the Map Reduce:

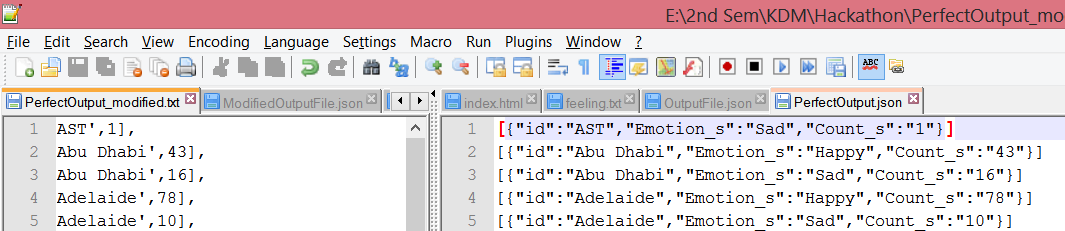




The Map Reduce actually can create different formats of the same output, by just tweeking the code a little.

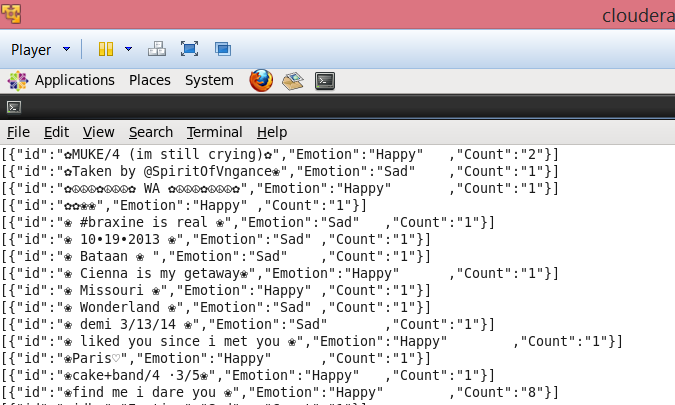
For example





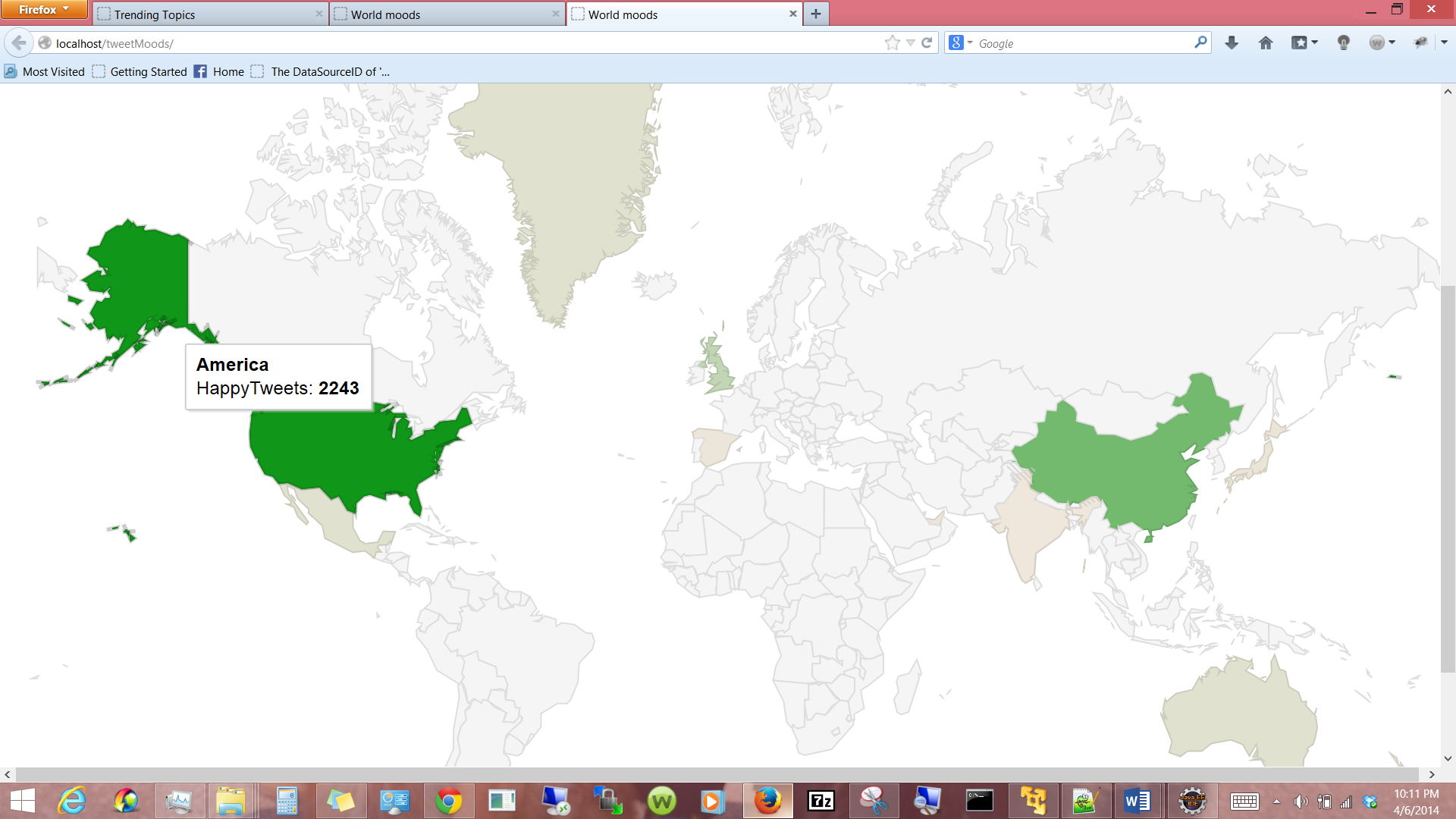
**Selection of data sets:** I have two primary data sets 1) The semantics of the tweet text 2) Location. I was able to pull both of these from the json response provided by twitter 1.1 rest API. For Location I had to use the “time\_zone” because when I used “user.location” object, since it was under user control, it was very inaccurate.

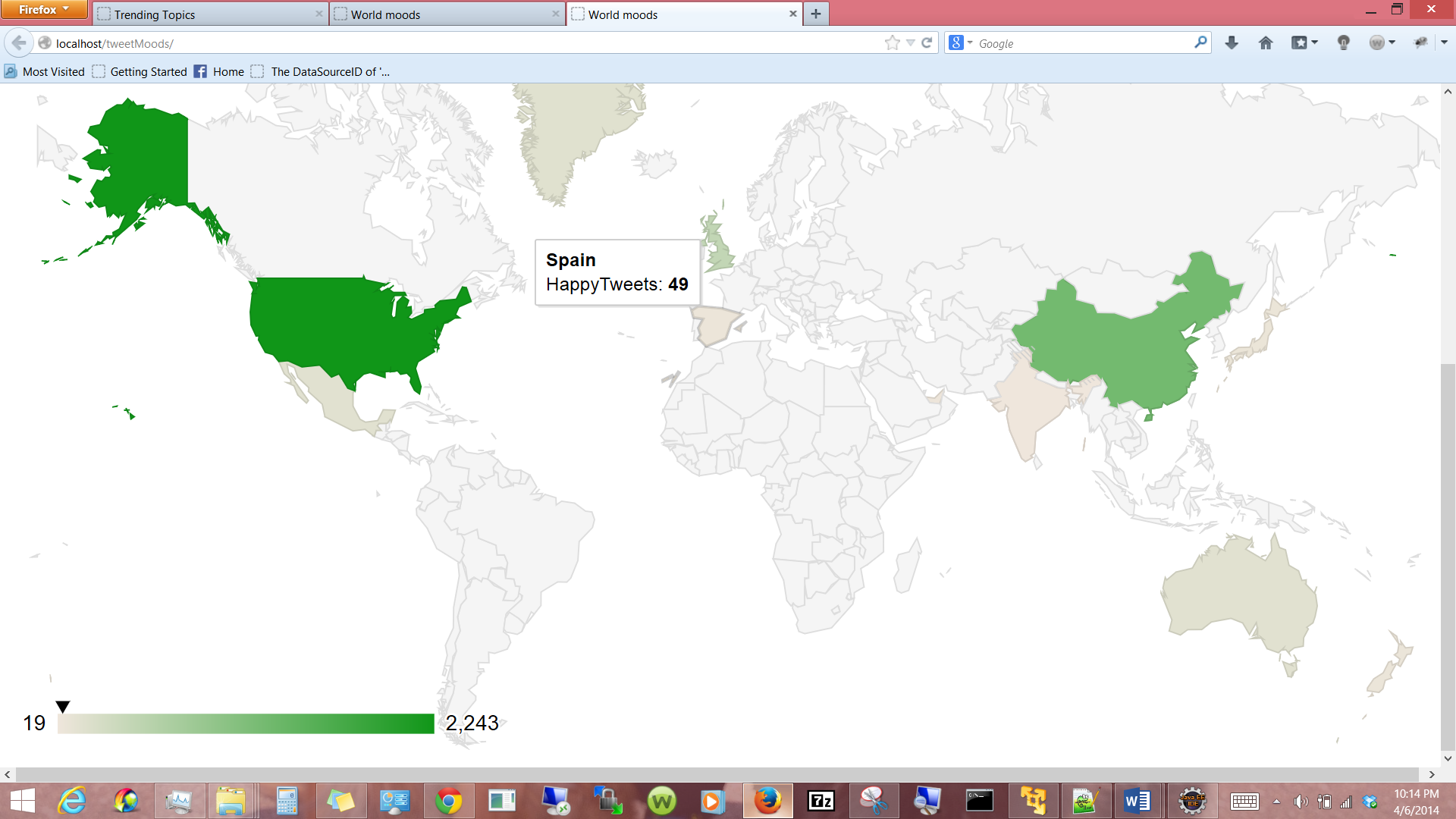
Example



**Web design:** Then I used the data to populate a google visualizations interactive webpage, the webpage contains a map which displays the number of happy tweets from different countries, we can see the how people in America and China are much happier than the rest of the world. China has pure numbers and America has a huge social media presence







**2. Features Implemented:**

1) Combined two datasets “location data” and “tweet text” To create a unique visualization.

2) Implemented my Own Classifier algorithm.

3) Combined Map Reduce, Classifer Algorithm and Json to text and text to Json converter into one single Map Reduce program.

**Software Services**

1. **My own services**

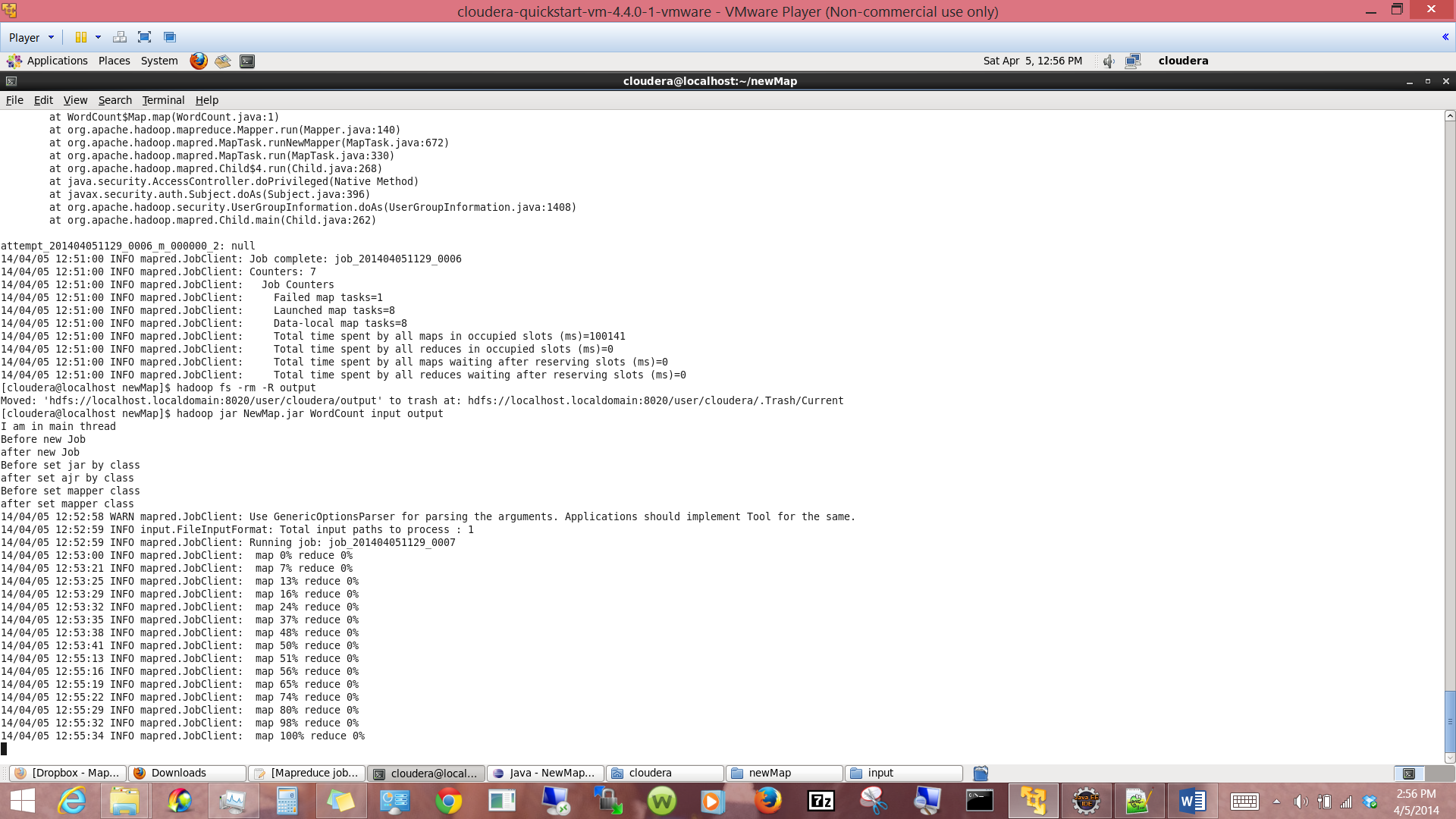
The all in one Map Reduce, Classifier, json converter.

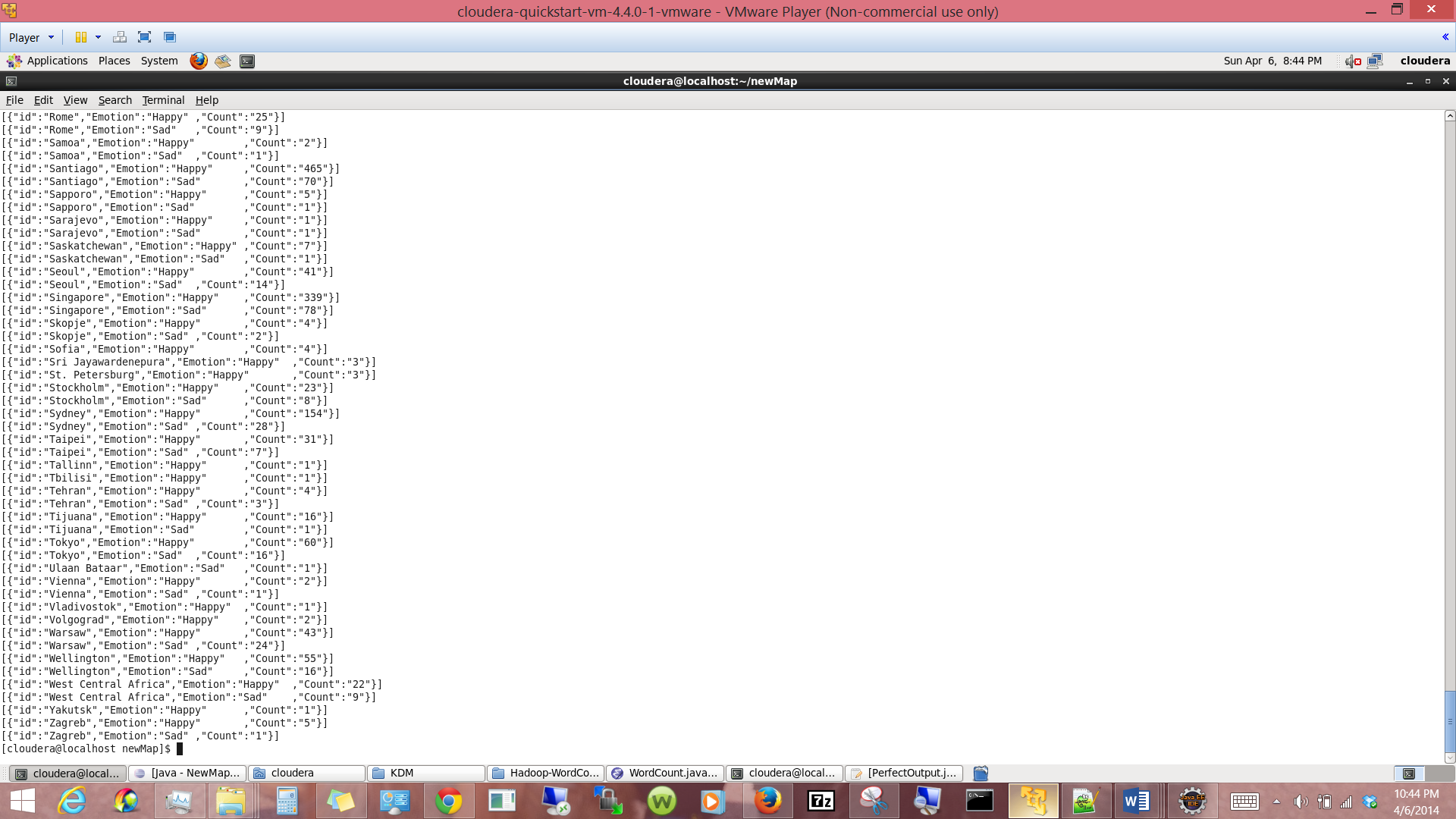
1. **Existing Services**

The Twitter Rest API and the Google visualization.

1. **Outputs: description with screenshots of the Features**

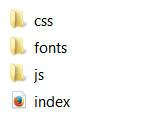
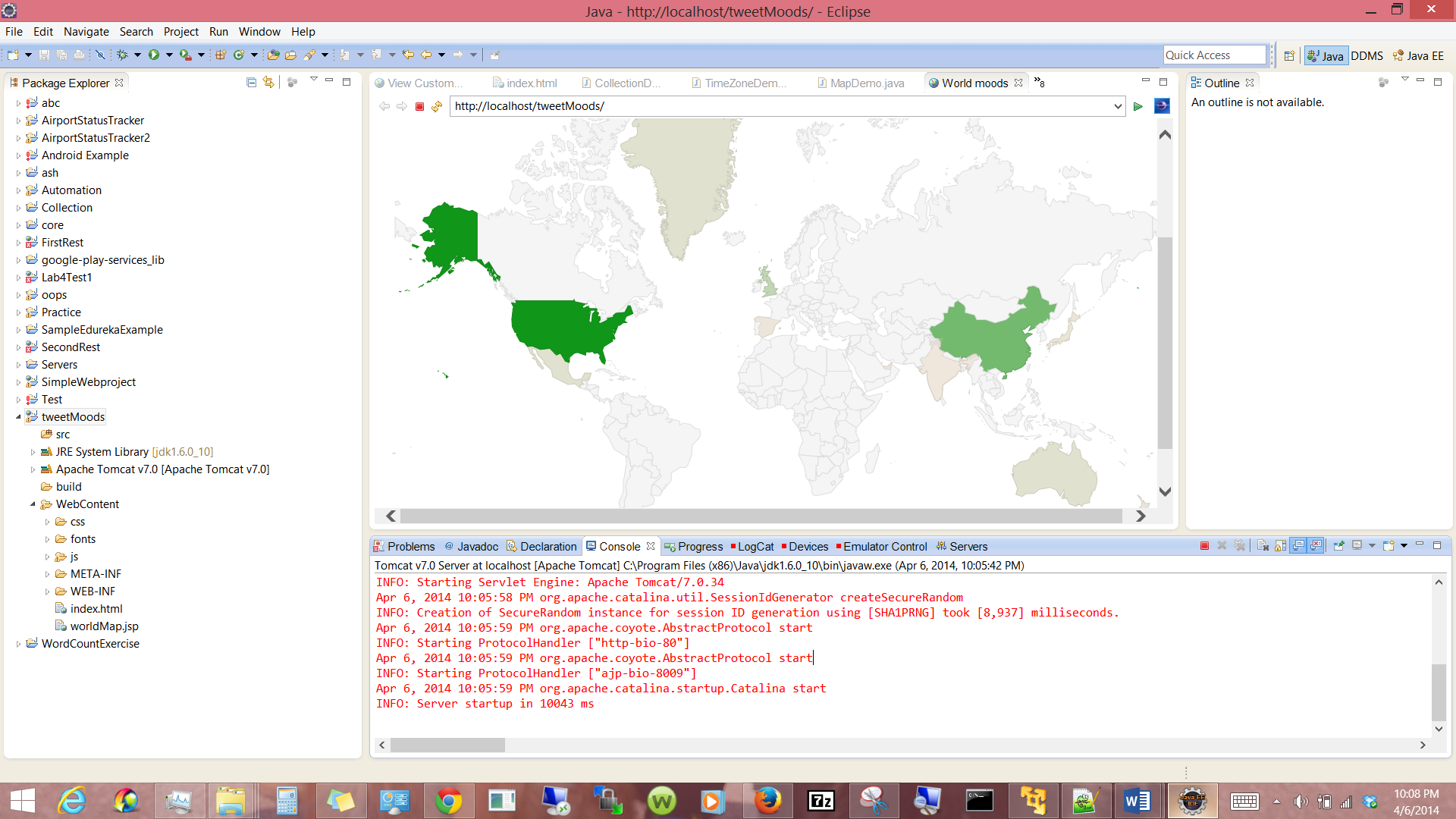
**Map Reduce:**

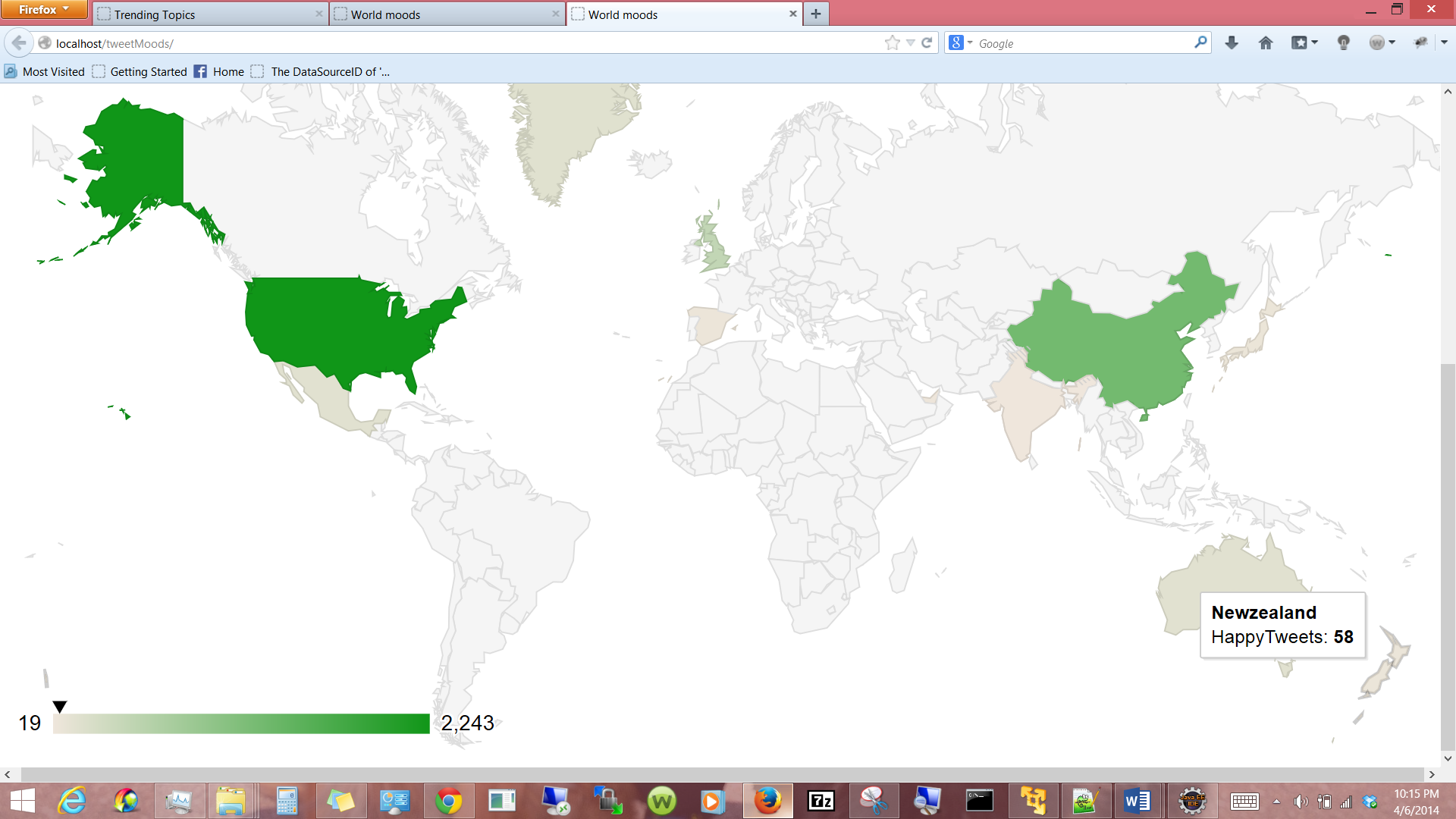




**Web Page:**

Inside the tweetMoods web project we can see the following structure.

**** 

As soon as the user runs the application on the server or the user double clicks on the “index” file the application is launched.

**4.All the Web Service and Web Site URLs**

Solr :

[http://134.193.136.127:8983/solr/new\_core/select?q=Emotion\_s%3A\*&rows=20&wt=json&indent=true](http://134.193.136.127:8983/solr/new_core/select?q=Emotion_s%3A*&rows=20&wt=json&indent=true)

**5. midterm Github URL**

**6. Limitations**

It’s hard to load all the countries in the single google visualization page. Could have created a much better and 100% dynamic web application if I had more time.

**7. References** <https://portal.futuregrid.org/manual/hadoop-wordcount>

<https://developers.google.com/chart/interactive/docs/gallery/geomap>

<https://github.com/uwescience/datasci_course_materials/blob/master/assignment1/twitterstream.py>

Materials Provided in class & tutorials