**Coursework Assignment**

**YUEFENG LIANG**

**201350099**

**Task**

The Libyan presidential elections are scheduled to take place on 10th December 2018 and the two candidates are Saif al-Islam Gaddafi (one of the sons of Muammar Gaddafi) and Khalifa Haftar (a general in the Libyan National Army). In order to predict the result of these elections, it is required to build a system to monitor Twitter and search tweets about the two candidates.

**Middleware configuration**

In this project, a storm local mode setup was used an Ubuntu machine on the Virtual machine. The reason why I choose the Storm local mode was because it is easy to test and debug one single Java process. Also, there is no need to use distributed mode to build the such easy system about monitor Twitter. The local storm mode was used to get the tweets which include the keywords, then the component of storm analyzed the result and predict the final winner.

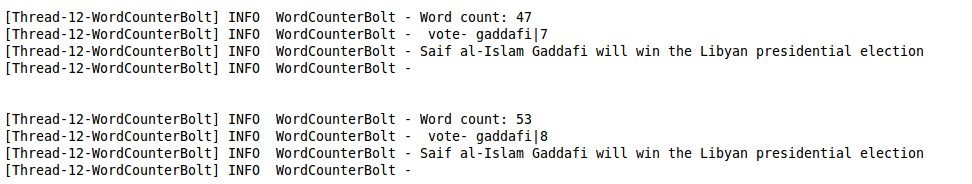
**Data Analytic Design**

Twitter is an online social network that provides a platform for people to send and receive the tweets. Firstly, it is required to build the spout to get the tweets as soon as possible. Twitter provides an API named “Twitter Streaming API”, which is a web-based service tool for retrieving tweets submitted by people in real time. Twitter4J is an unofficial Java library for the Twitter API. With Twitter4J, it is easy to integrate the Java application with the Twitter service. However, the Virtual machine needs to get the authentication details (consumerKey, consumerSecret, accessToken, accessTokenSecret) from Twitter before using Twitter Streaming API. Then, a TwittterSampleSpout was built to search the tweets (see Appendix1). Furthermore, the tweets sent by the spout will be transferred to WordSplitterBolt (see Appendix2). Bolt is a component that takes tuples as input, processes tuples, and generates new tuples as output. Thus, the WordSplitterBolt would split the tweets to each word and generate a new tuple. Another bolt, name “WordCounterBolt” (see Appendix3), would use the new tuple to calculate the number of times each word appears and output the number of times about surname of two candidates appeared in tweets. According these outputs, it is convenient to predict who will win the Libyan presidential elections. Finally, submitting the topology is the main application (see Appendix4). The spout, two bolts are all initialized in the topology.

In a word, the complete application has five Java code, which are topology.java, TwittterSampleSpout.java, WordSplitterBolt.java, WordCounterBolt.java and IgnoreWordsBolt.java. The IgnoreWordsBolt is just used to ignore the normal words (see Appendix5).

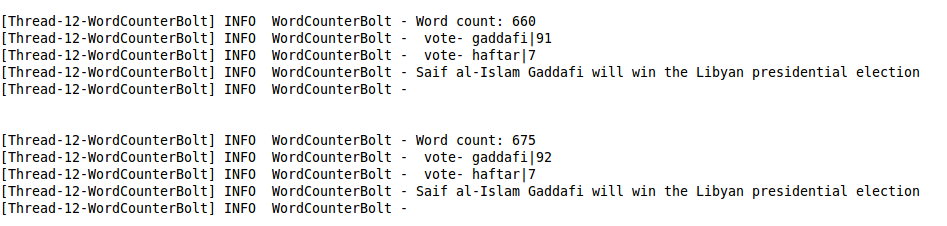
**Result:**

According to the WordCounterBolt.java code, the bolt outputs the data to the logger every 10 seconds. At the beginning, the system only found the tweets about “Gaddafi”, and the word count number mean the numbers of words in tweets which included the word “Gaddafi”. So, Figure 1 shows the output of the system in first 20 minutes.

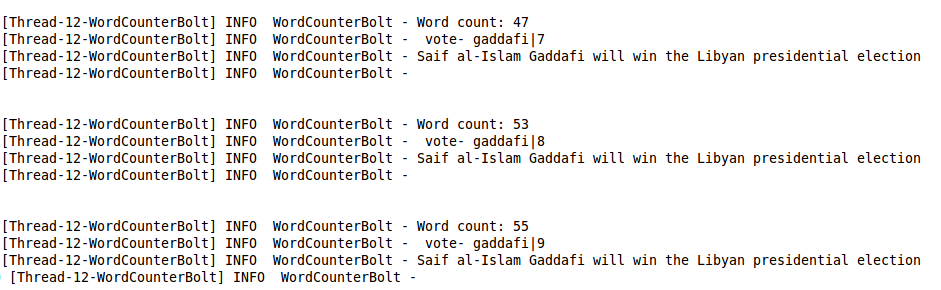


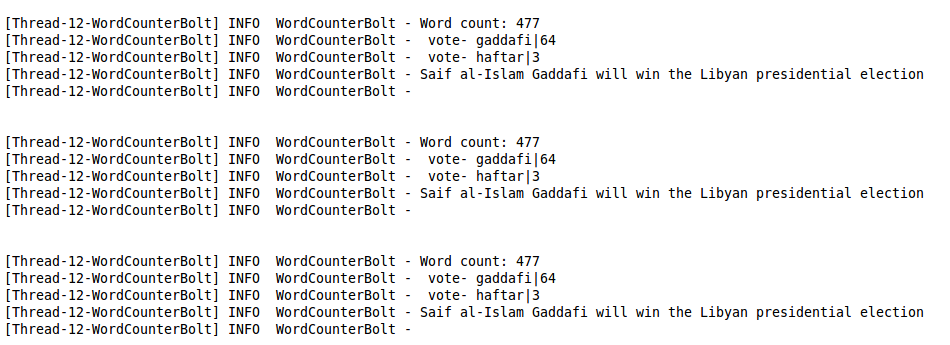
**Figure 1. the data of the storm Twitter project in the first 20 minutes**

Then, after few hours, first tweet with “Haftar” showed up. In this project, it cost much time to search the tweets with the word “Haftar”, which means the tweets with “Haftar” were less than the tweets with “Gaddafi”. Figure 2 shows that the output of the java application after few hours.



**Figure 2. the data of the storm Twitter project after many hours**

****Sometimes, system produced the same data in two or three outputs. It may be due to TwittterSampleSpout found some wrong data or the spout could not find the tweets with either the word “Gaddafi” or “Haftar”. Figure3 shows the same data produced by the system.



**Figure 3. the wrong data output from the system**

**Discussion of Results:**

According to the data produced by the system after many hours, it could be predicted that Saif al-Islam Gaddafi may win the Libyan presidential elections, as the numbers of times of tweets with the word “Gaddafi” appeared more than that with “Haftar”.

**Conclusions and Recommendations**

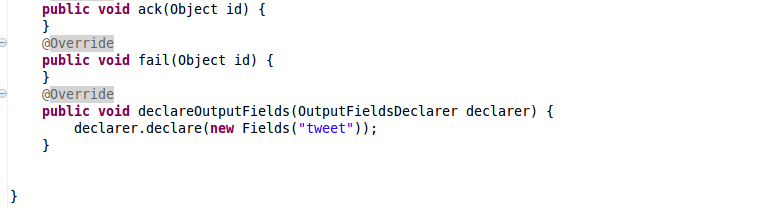
It is not difficult to find out the tweets with the names of two candidates. However, it is a simple but not reliable method to predict the presidential elections. During the process, it is difficult to identify if these tweets were voting for the candidates. It is necessary to analyze each tweet in detail, which may cost much time.

Appendix1

Appendix1 – TwittterSampleSpout.java was used to search the tweets.







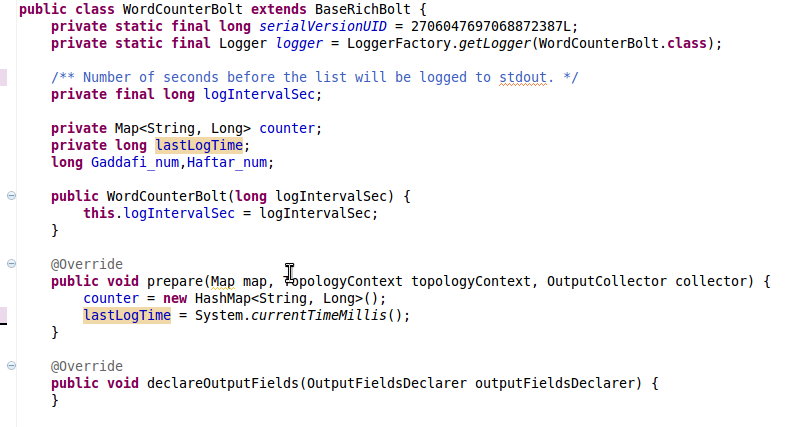
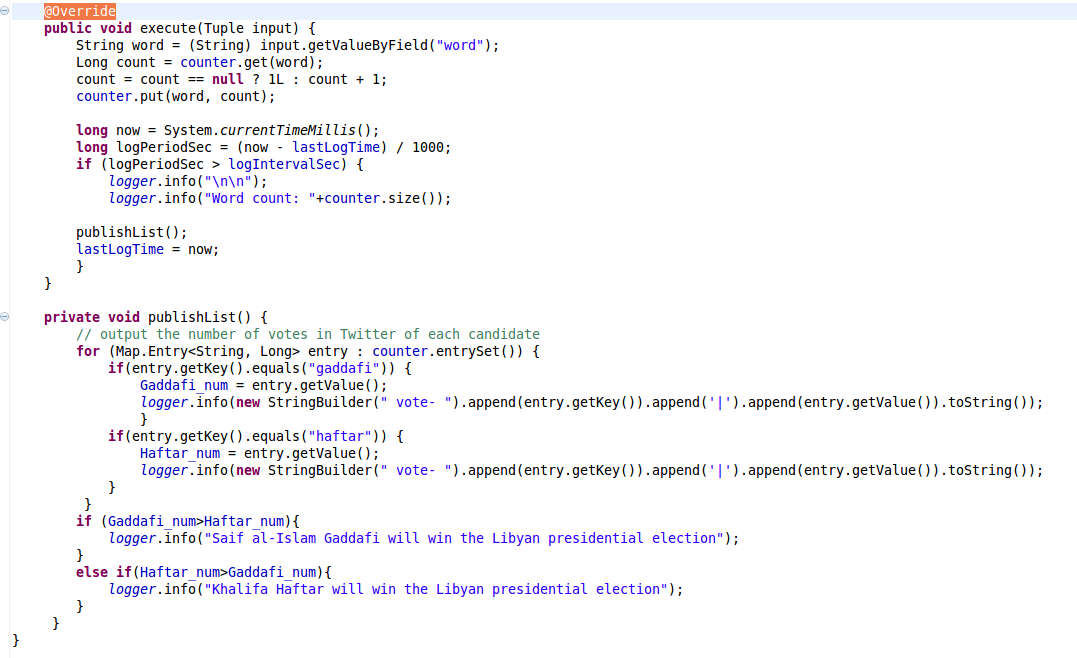
Appendix2

Appendix2 – WordSplitterBolt.java would split the tweets to each word and generate a new tuple.



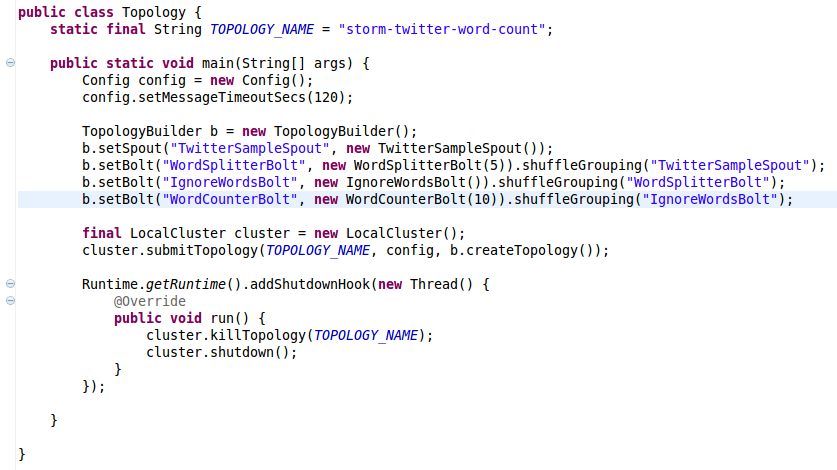
Appendix3

Appendix3 – WordCounterBolt.java calculate and output the results.



Appendix4

Appendix4 –Topology.java initializes the spout and three bolts.



Appendix5

Appendix5 – IgnoreWordsBolt.java used to ignore the normal words

