1. **Word count on tweets :**

* Input: Tweets for a given domain( Cricket-IPL-2017) - stored in tweets.txt
* Processing: The tweets.txt file was fed to MR on HDFS for processing of the word count
* Output: The output of that was transformed to a excel format and fed to R for Word-cloud. Tweets\_cricket.xlsx.
* JAR file : wc\_1.jar
* Output folder : count\_tw\_1 and R notebook

1. **Word co-occurrence on tweets :**

* Used two techniques Word Pair and Word Stripes :
* Input : tweets.txt file used in the above activity
* Processing : Writing a MR program
* Output : Co-occurrence pairs and stripes
* JAR file for Pairs : wdp.jar
* JAR file for Stripes : ws\_final.jar
* Output folder for Pairs : out\_2\_par
* Output folder for Stripes : out\_tw\_final1

1. **Featured Activity 1: Word count on Classical Latin text:**

* Input : The two latin files given along with the lemmatization.csv file
* Processing: Cleaned the data inside the Mapper and the did the processing of MR on HDFS.
* Output : word <location\_1><location\_2>...<location\_n>

lemma<location\_1><location\_2>...<location\_n>

* Jar File : count.jar
* Output : out\_count\_4

**4. Featured Activity 2: Word co-occurrence among multiple documents.**

* Graph plotted using the same code as used in activity 2.
* Output: The R notebook is given

**References Used :**

1. <https://dzone.com/articles/calculating-co-occurrence>
2. <http://stackoverflow.com/questions/23209174/converting-mapwritable-to-a-string-in-hadoop>
3. <https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>