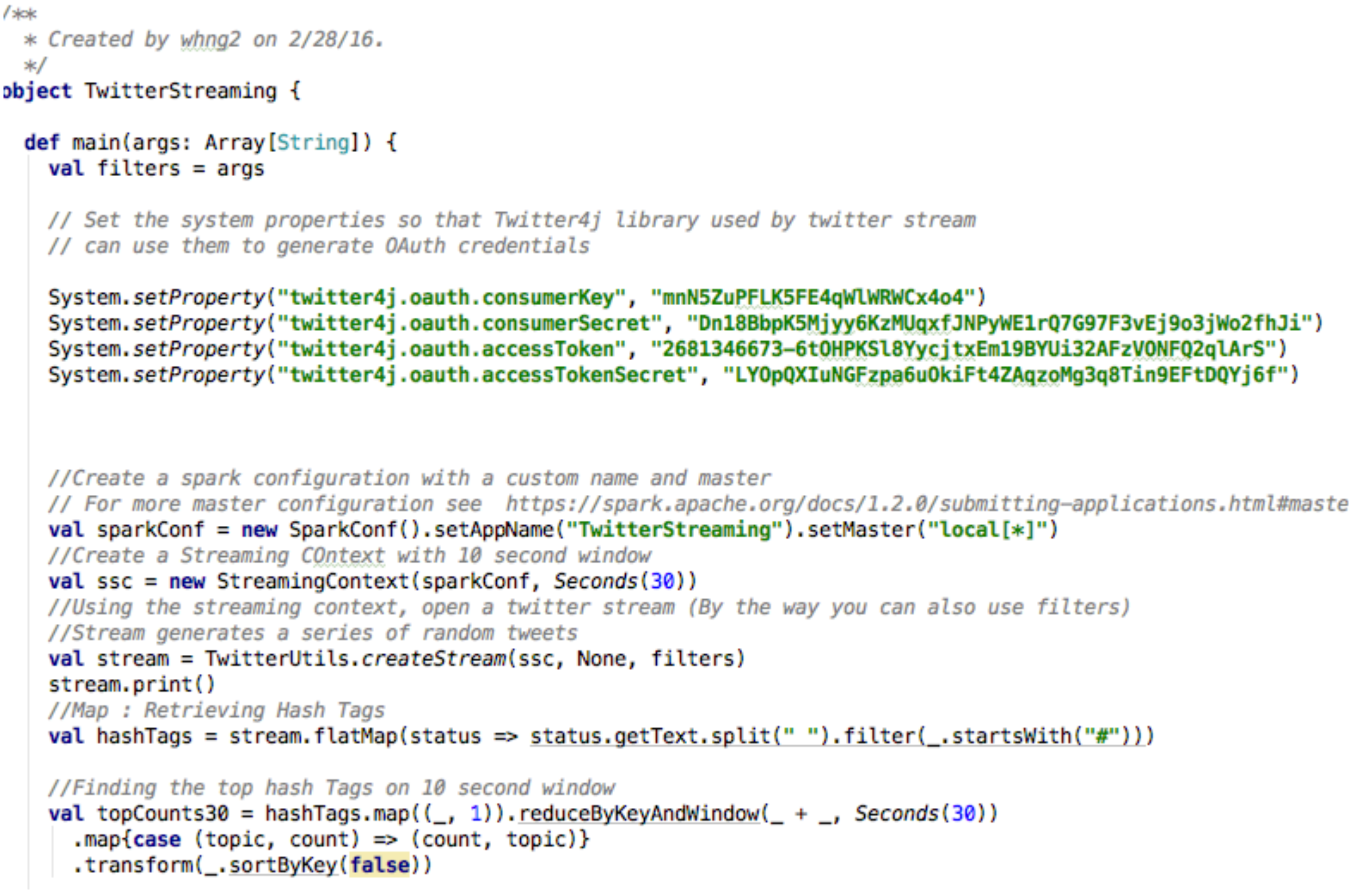
1.
Question:

Implement a smart application with big data analytics related to your project showing the collaboration between Spark and Smart Apps. Implement Twitter Streaming and perform word count on it and publish the results and showcase it in your Smart Phone/Watch Application.

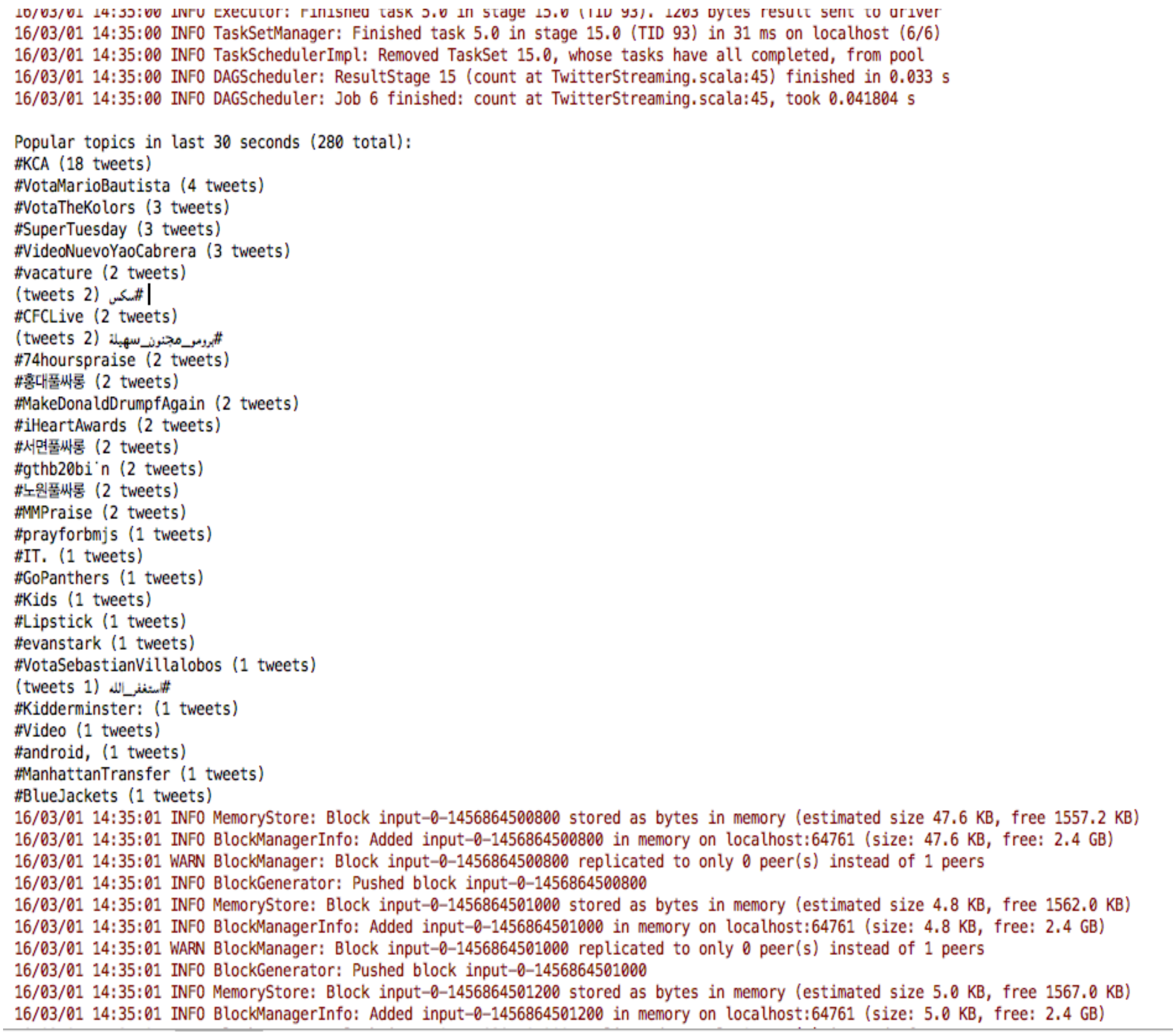
Description:

Based on Lab 5 twitter streaming lab example source code, connect with Twitter account and sort the popular tweeting words for certain period of time and then do word count.

Screenshot:



```
/**
 * Created by whng2 on 2/28/16.
 */
object TwitterStreaming {

  def main(args: Array[String]) {
    val filters = args

    // Set the system properties so that Twitter4j library used by twitter stream
    // can use them to generate OAuth credentials

    System.setProperty("twitter4j.oauth.consumerKey", "mnN5ZuPFLK5FE4qWlWRWCx4o4")
    System.setProperty("twitter4j.oauth.consumerSecret", "Dn18BbpK5Mjyy6KzMUqxfJNPyWE1rQ7G97F3vEj9o3jWo2fhJi")
    System.setProperty("twitter4j.oauth.accessToken", "2681346673-6tOHPKSl8YycjtxEm19BYUi32AFzVONFQ2qlArS")
    System.setProperty("twitter4j.oauth.accessTokenSecret", "LYOpQXIuNGFzpa6uOkiFt4ZAqzoMg3q8Tin9EFtDQYj6f")



    //Create a spark configuration with a custom name and master
    // For more master configuration see  https://spark.apache.org/docs/1.2.0/submitting-applications.html#maste
    val sparkConf = new SparkConf().setAppName("TwitterStreaming").setMaster("local[*]")
    //Create a Streaming COntext with 10 second window
    val ssc = new StreamingContext(sparkConf, Seconds(30))
    //Using the streaming context, open a twitter stream (By the way you can also use filters)
    //Stream generates a series of random tweets
    val stream = TwitterUtils.createStream(ssc, None, filters)
    stream.print()
    //Map : Retrieving Hash Tags
    val hashTags = stream.flatMap(status => status.getText.split(" ").filter(_.startsWith("#")))

    //Finding the top hash Tags on 10 second window
    val topCounts30 = hashTags.map((_, 1)).reduceByKeyAndWindow(_ + _, Seconds(30))
      .map{case (topic, count) => (count, topic)}
      .transform(_.sortByKey(false))
```



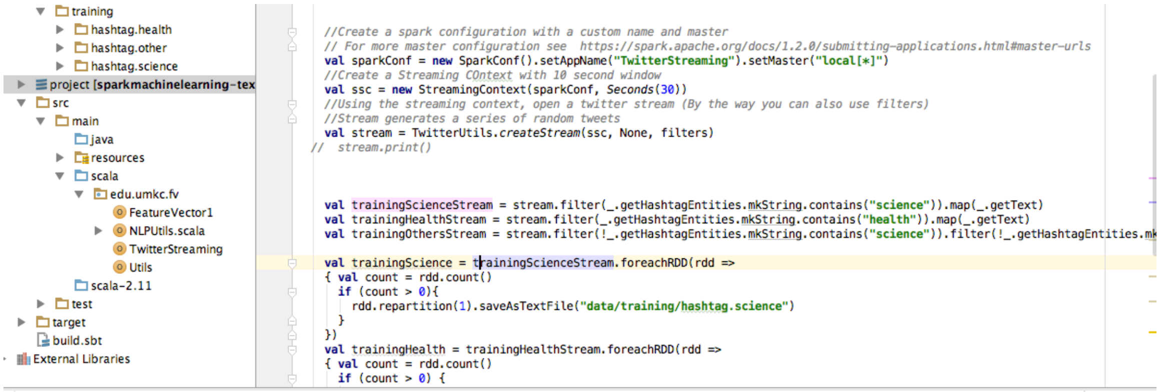
```
16/03/01 14:35:00 INFO Executor: Finished task 5.0 in stage 15.0 (TID 93). 1203 bytes result sent to driver
16/03/01 14:35:00 INFO TaskSetManager: Finished task 5.0 in stage 15.0 (TID 93) in 31 ms on localhost (6/6)
16/03/01 14:35:00 INFO TaskSchedulerImpl: Removed TaskSet 15.0, whose tasks have all completed, from pool
16/03/01 14:35:00 INFO DAGScheduler: ResultStage 15 (count at TwitterStreaming.scala:45) finished in 0.033 s
16/03/01 14:35:00 INFO DAGScheduler: Job 6 finished: count at TwitterStreaming.scala:45, took 0.041804 s

Popular topics in last 30 seconds (280 total):
#KCA (18 tweets)
#VotaMarioBautista (4 tweets)
#VotaTheKolors (3 tweets)
#SuperTuesday (3 tweets)
#VideoNuevoYaoCabrera (3 tweets)
#vacature (2 tweets)
(tweets 2) سكس#
#CFCLive (2 tweets)
(tweets 2) برومو_مجنون_سهيلة#
#74hourspraise (2 tweets)
#홍대풀싸롱 (2 tweets)
#MakeDonaldDrumpfAgain (2 tweets)
#iHeartAwards (2 tweets)
#서면풀싸롱 (2 tweets)
#gthb20bi˙n (2 tweets)
#노원풀싸롱 (2 tweets)
#MMPraise (2 tweets)
#prayforbmjs (1 tweets)
#IT. (1 tweets)
#GoPanthers (1 tweets)
#Kids (1 tweets)
#Lipstick (1 tweets)
#evanstark (1 tweets)
#VotaSebastianVillalobos (1 tweets)
(tweets 1) استغفر_الله#
#Kidderminster: (1 tweets)
#Video (1 tweets)
#android, (1 tweets)
#ManhattanTransfer (1 tweets)
#BlueJackets (1 tweets)
16/03/01 14:35:01 INFO MemoryStore: Block input-0-1456864500800 stored as bytes in memory (estimated size 47.6 KB, free 1557.2 KB)
16/03/01 14:35:01 INFO BlockManagerInfo: Added input-0-1456864500800 in memory on localhost:64761 (size: 47.6 KB, free: 2.4 GB)
16/03/01 14:35:01 WARN BlockManager: Block input-0-1456864500800 replicated to only 0 peer(s) instead of 1 peers
16/03/01 14:35:01 INFO BlockGenerator: Pushed block input-0-1456864500800
16/03/01 14:35:01 INFO MemoryStore: Block input-0-1456864501000 stored as bytes in memory (estimated size 4.8 KB, free 1562.0 KB)
16/03/01 14:35:01 INFO BlockManagerInfo: Added input-0-1456864501000 in memory on localhost:64761 (size: 4.8 KB, free: 2.4 GB)
16/03/01 14:35:01 WARN BlockManager: Block input-0-1456864501000 replicated to only 0 peer(s) instead of 1 peers
16/03/01 14:35:01 INFO BlockGenerator: Pushed block input-0-1456864501000
16/03/01 14:35:01 INFO MemoryStore: Block input-0-1456864501200 stored as bytes in memory (estimated size 5.0 KB, free 1567.0 KB)
16/03/01 14:35:01 INFO BlockManagerInfo: Added input-0-1456864501200 in memory on localhost:64761 (size: 5.0 KB, free: 2.4 GB)
```

2. Question:

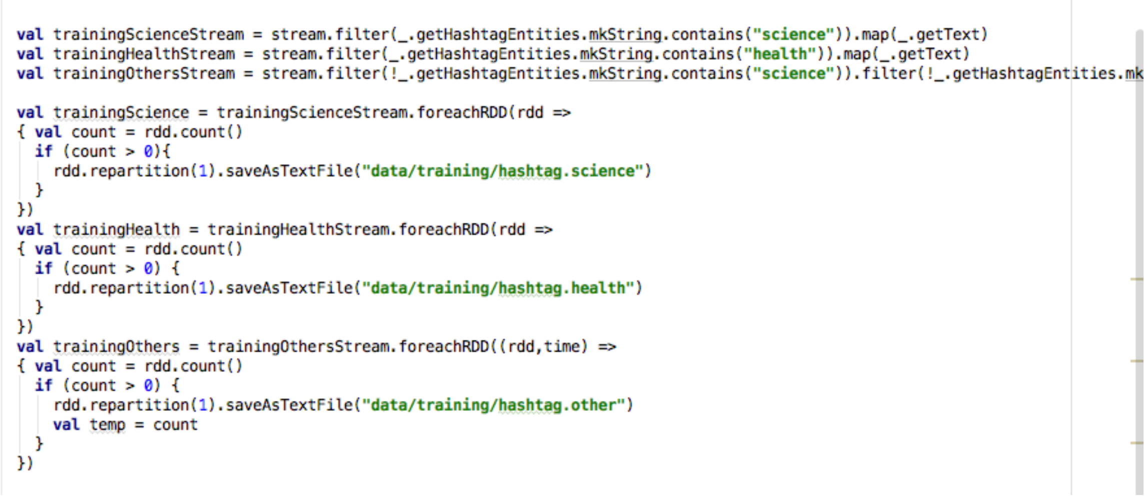
Perform a machine-learning algorithm with the Twitter Streaming data to categorize each Tweet
1) Training datasets: Collect different categories of Tweets related to your project. (Categories can be based on HashTags / Subjects etc.)
2) Test data: the upcoming twitter stream.

Description:
Using Twitter streaming source code to categorize twitter streaming messages as Science, health and others. Then based on the training data to test coming test data to see if the message belongs to science, health, or just others.

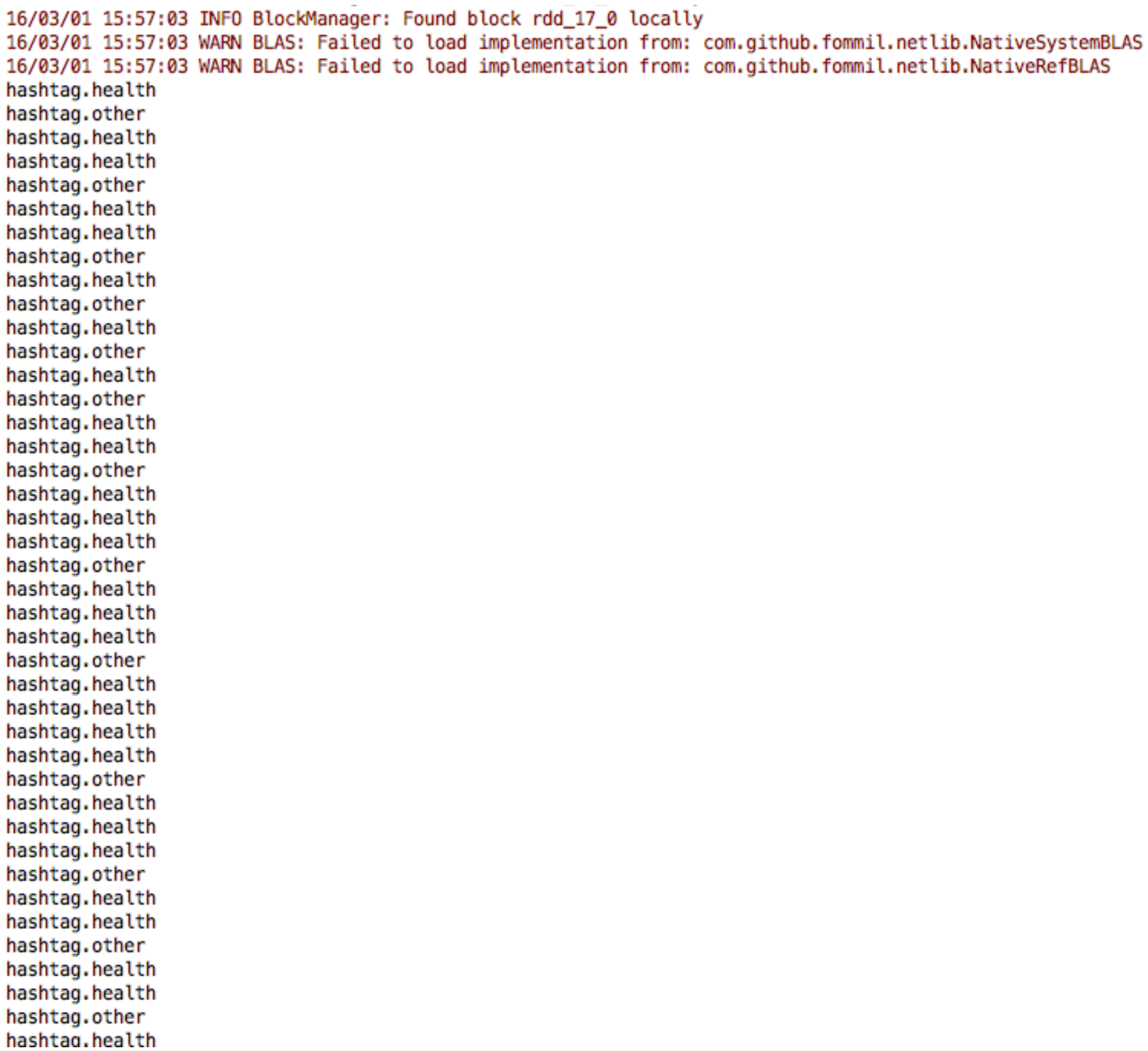
Screen shot:



```
  //Create a spark configuration with a custom name and master
  // For more master configuration see  https://spark.apache.org/docs/1.2.0/submitting-applications.html#master-urls
  val sparkConf = new SparkConf().setAppName("TwitterStreaming").setMaster("local[*]")
  //Create a Streaming COntext with 10 second window
  val ssc = new StreamingContext(sparkConf, Seconds(30))
  //Using the streaming context, open a twitter stream (By the way you can also use filters)
  //Stream generates a series of random tweets
  val stream = TwitterUtils.createStream(ssc, None, filters)
//  stream.print()



  val trainingScienceStream = stream.filter(_.getHashtagEntities.mkString.contains("science")).map(_.getText)
  val trainingHealthStream = stream.filter(_.getHashtagEntities.mkString.contains("health")).map(_.getText)
  val trainingOthersStream = stream.filter(!_.getHashtagEntities.mkString.contains("science")).filter(!_.getHashtagEntities.mk

  val trainingScience = trainingScienceStream.foreachRDD(rdd =>
  { val count = rdd.count()
    if (count > 0){
      rdd.repartition(1).saveAsTextFile("data/training/hashtag.science")
    }
  })
  val trainingHealth = trainingHealthStream.foreachRDD(rdd =>
  { val count = rdd.count()
    if (count > 0) {
```

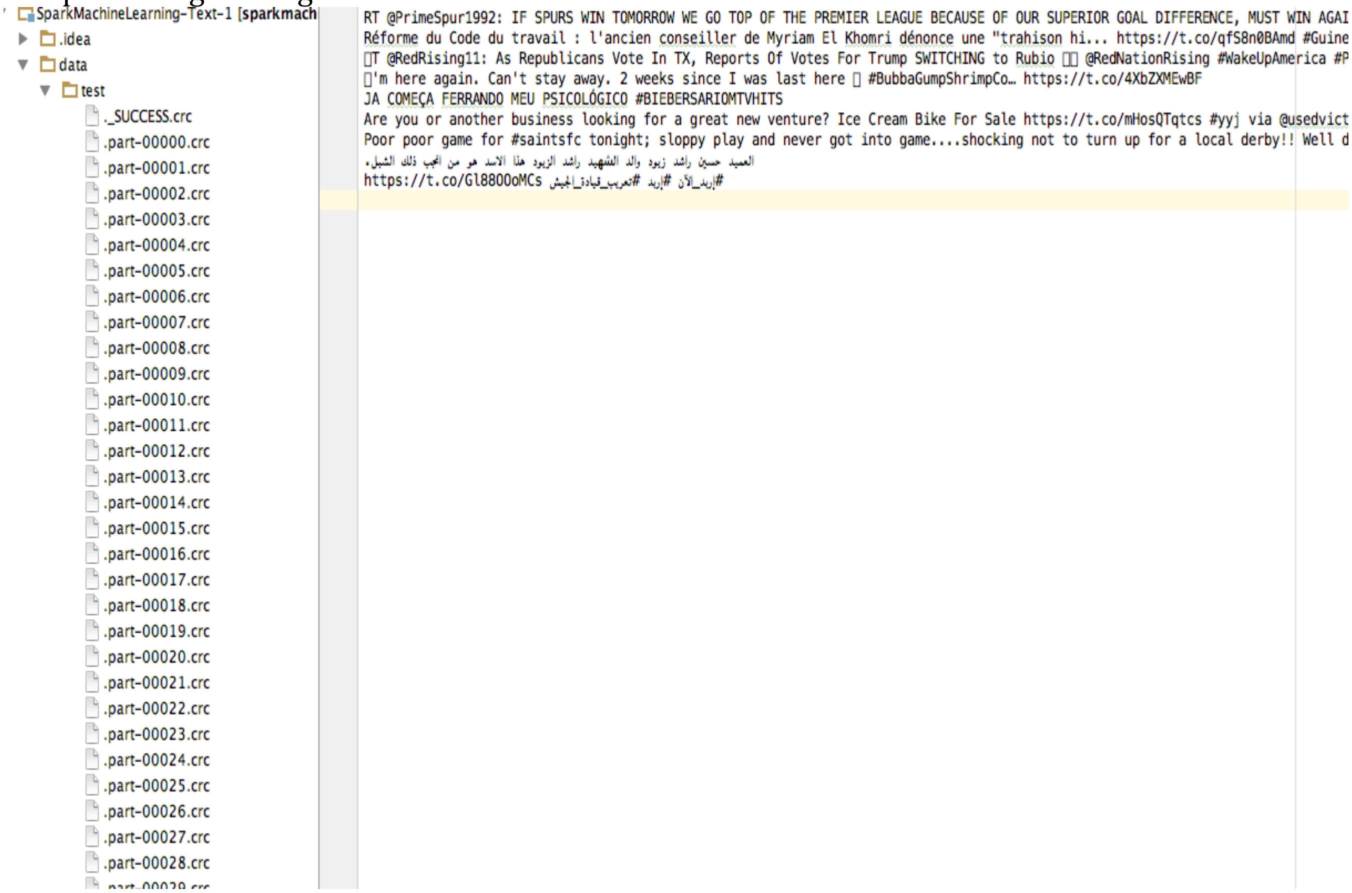


```
val trainingScienceStream = stream.filter(_.getHashtagEntities.mkString.contains("science")).map(_.getText)
val trainingHealthStream = stream.filter(_.getHashtagEntities.mkString.contains("health")).map(_.getText)
val trainingOthersStream = stream.filter(!_.getHashtagEntities.mkString.contains("science")).filter(!_.getHashtagEntities.mk

val trainingScience = trainingScienceStream.foreachRDD(rdd =>
{ val count = rdd.count()
  if (count > 0){
    rdd.repartition(1).saveAsTextFile("data/training/hashtag.science")
  }
})
val trainingHealth = trainingHealthStream.foreachRDD(rdd =>
{ val count = rdd.count()
  if (count > 0) {
    rdd.repartition(1).saveAsTextFile("data/training/hashtag.health")
  }
})
val trainingOthers = trainingOthersStream.foreachRDD((rdd,time) =>
{ val count = rdd.count()
  if (count > 0) {
    rdd.repartition(1).saveAsTextFile("data/training/hashtag.other")
    val temp = count
  }
})
```



```
16/03/01 15:57:03 INFO BlockManager: Found block rdd_17_0 locally
16/03/01 15:57:03 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
16/03/01 15:57:03 WARN BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.other
hashtag.health
hashtag.other
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.other
hashtag.health
hashtag.health
hashtag.other
hashtag.health
```

Sample testing message received from twitter account:



```
RT @PrimeSpur1992: IF SPURS WIN TOMORROW WE GO TOP OF THE PREMIER LEAGUE BECAUSE OF OUR SUPERIOR GOAL DIFFERENCE, MUST WIN AGAI
Réforme du Code du travail : l'ancien conseiller de Myriam El Khomri dénonce une "trahison hi... https://t.co/qfS8n0BAmd #Guine
□T @RedRising11: As Republicans Vote In TX, Reports Of Votes For Trump SWITCHING to Rubio □□ @RedNationRising #WakeUpAmerica #P
□'m here again. Can't stay away. 2 weeks since I was last here □ #BubbaGumpShrimpCo… https://t.co/4XbZXMEwBF
JA COMEÇA FERRANDO MEU PSICOLÓGICO #BIEBERSARIOMTVHITS
Are you or another business looking for a great new venture? Ice Cream Bike For Sale https://t.co/mHosQTqtcs #yyj via @usedvict
Poor poor game for #saintsfc tonight; sloppy play and never got into game....shocking not to turn up for a local derby!! Well d
العميد حسين راشد زيود والد الشهيد راشد الزيود هذا الاسد هو من انجب ذلك الشبل.
https://t.co/Gl88OOoMCs #تعريب_قيادة_الجيش #اربد #اربد_الآن
```