Part 1

Download Twitter Data to MySQL

Kimya Dhade Kalyani Deshmukh Sylvia Shun

```
package twitterExtractor;
//All the required Imports
import com.cybozu.labs.langdetect.Detector;
import com.cybozu.labs.langdetect.DetectorFactory;

import twitter4j.*;

import weka.core.Instances;
import weka.core.converters.ArffSaver;

import java.io.BufferedWriter;
import java.io.File;
import java.io.FileWriter;

import java.sql.Connection;
import java.sql.Statement;

import java.util.*;
import java.util.regex.Matcher;
import java.util.regex.Pattern;

public class TweetListener2 implements StatusListener {

   private double numTotalTweets;
   private Connection conn;
   private HashSet<Long> users;
   private static Detector detector;

   public TweetListener2(Connection conn1) {
      conn = conn1;
      numTotalTweets = 0.0;
      users = new HashSet<Long>();

     try {
         DetectorFactory.loadProfile("profiles");
```

```
    } catch (Exception e) {
        System.out.println("Ciao");
        e.printStackTrace();  //To change body of catch statement use File | Settings | File
Templates.
        System.exit(1);
    }

  }

  public void onException(java.lang.Exception ex) {
    ex.printStackTrace();
    System.exit(1);
  }

  public void onTrackLimitationNotice(int numberOfLimitedStatuses) {
    try {
      System.out.println("----SOSPENSION---");
      Thread.sleep(3600000);
    } catch (Exception e) {
      e.printStackTrace();
      System.exit(1);
    }
  }

  public void onScrubGeo(long l, long l1) {
    //To change body of implemented methods use File | Settings | File Templates.
  }

  public void onStallWarning(StallWarning t) {
    try {

    } catch (Exception e) {
      e.printStackTrace();
      System.exit(1);
    }}

  public void onDeletionNotice(StatusDeletionNotice status) {
  }

  public void onStatus(Status status) {
    try {

      //Thread.sleep(30000);
```

```
        //System.out.println("");
        // JSONObject json = new JSONObject(status);
        String tweet = status.getText();
        tweet = tweet.replaceAll("\\,", " ");
        tweet = tweet.replaceAll("\"", " ");
//To handle special characters due to Emoticons used in Tweets
        tweet = tweet.replaceAll("\\p{C}", "");

        //tweet = tweet.replaceAll("#","\\#");
        // System.out.println("CULO"+tweet);

        Date dateTweet = status.getCreatedAt();

        User user = status.getUser();
        long idtweet = status.getId();
        long iduser = user.getId();
        Date userCreateAt = user.getCreatedAt();
        int numFollowers = user.getFollowersCount();
        int numFriends = user.getFriendsCount();


        java.text.SimpleDateFormat sdf = new java.text.SimpleDateFormat("yyyy-MM-dd
HH:mm:ss");
        String dateTweet1 = sdf.format(dateTweet);
        String userCreatedAt1 = sdf.format(userCreateAt);

        detector = DetectorFactory.create();
        detector.append(tweet);
        String lang = detector.detect();
        String patternStr = "(?:\\s|\\A)[##]+([A-Za-z0-9-_]+)" ;
            Pattern pattern = Pattern.compile(patternStr);
            Matcher matcher = pattern.matcher(tweet);
            String result = "";

  // Search for Hashtags
            while (matcher.find()) {
               result = matcher.group();
               result = result.replace(" ", "");
               String search = result.replace("#", "");
               Statement stmt = conn.createStatement();
               System.out.println(" "+result);
//Extract the Hashtag from the Tweet for trend analysis
              stmt.executeUpdate("insert into twitter.hashtag (hashtag) value ('" + search +" ')");
```

```
        stmt.close();
              }

         if (lang.equals("en")) {
//Export tweets data to tweet table in MySQL
            Statement stmt = conn.createStatement();
            String query = "insert into twitter.tweet (idTweet,text,date,iduser) value (" + idtweet +
",\"" + tweet.trim() + "\",\"" + dateTweet1 + "\"," + iduser + ")";
             System.out.println("Saved in array");
            stmt.executeUpdate(query);
            stmt.close();


            if (!users.contains(iduser)) {
//Export User data to user table in MySQL
              Statement stmt1 = conn.createStatement();
              String query1 = "insert into twitter.user (iduser, numFollowers, numFriends,
createdAt) value (" + iduser + "," + numFollowers + "," + numFriends + ",\"" + userCreatedAt1 +
"\")";
              stmt1.executeUpdate(query1);
              stmt1.close();
              users.add(iduser);

            }


            numTotalTweets++;
            if (numTotalTweets % 500 == 0) {
              System.out.println("Numero Tweet: " + numTotalTweets);
            }
            if (numTotalTweets == 1000000) {
              System.exit(0);
            }
          }


          System.out.println(dateTweet);


          //UseRecordsFile.writeAuthorInfo(iduser, userCreateAt, numFollowers, numFriends);
          //UseRecordsFile.writeTweetAuthor(idtweet, iduser);
          //UseRecordsFile.writeTweetDate(idtweet, dateTweet);
//            interfaceDb.set("tweet-text", idtweet, tweet);
//            interfaceDb.set("user-date", iduser, userCreateAt.toString());
```

```
//              interfaceDb.set("user-numFollowers", iduser, numFollowers);
//              interfaceDb.set("user-numFriends", iduser, numFriends);
//              interfaceDb.set("tweet-author", idtweet, iduser);
//              interfaceDb.set("tweet-date", idtweet, dateTweet.toString());


          //  System.out.println("onStatus got: " + status);
        } catch (Exception e) {

          e.printStackTrace();

        }
    }

}
```