Realtime Twitter Sentiment Analytics

Objective: To implement/ practice Apache spark features on a realtime application.
This application collects the realtime streaming tweets from twitter for target keywords and process the data in realtime.

Tools Used: Apache spark SCALA (Core API), Apache spark SQL, Apache spark Streaming, Twitter4j

Realtime Analytics performed:
- Tweet Sentiment of each tweet using lexical affinity method.
- Popular Retweets
- Top 5 hashtags
- Least 5 hashtags

Reusability:
This has been implemented in a generic way.
- Application code can be tweaked to process any realtime data from any HDFS/ local file system location.
  Eg: Log files
- Complex twitter data analytics can be performed by extending this code.

Observation:
- Results from this tweet analyzer are similar to the Rstudio's twitter analysis application.
  http://spark.rstudio.com/mczerny/TwitterAnalysis/
- Observed that most of the tweets appeared at almost same-time in twitter streaming and in the tweet analyzer application.
  https://twitter.com/search?f=realtime