twitteR_Intro

Viraj @DataTasteMaker

March 19, 2016

This document introduces the basic usage of twitteR package. In this document we will connect to the twitter and try to read the tweets of a particular word **"DataScience"**.

Also we will read the timeline of dataTasteMaker and try to find out what is he up to.

Before starting with this code, we need to have :

1. Twitter account
2. Created Twitter App
3. The below 4 keys
    - TWITTER_CONSUMER_KEY
    - TWITTER_CONSUMER_SECRET
    - TWITTER_ACCESS_TOKEN
    - TWITTER_ACCESS_TOKEN_SECRET

Clear the Environment and plots, to free up the memory

Load the required libraries, ignore the output in my script, you should simply note that the libraries are loaded properly

- twitteR
- wordcloud
- dplyr
- stringr

```
library(twitteR)    # for accessing and reading the tweets
library(RColorBrewer) # dependent library for wordcloud (may not be required for all, however my progra
library(wordcloud)  # for creating the word cloud of the tweets
library(stringr)    # for string clean up
library(dplyr)    # for formatting of data frame, selecting etc.
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:twitteR':
##
##     id, location

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Load the keys

```
TWITTER_CONSUMER_KEY = Sys.getenv("TWITTER_CONSUMER_KEY")
TWITTER_CONSUMER_SECRET = Sys.getenv("TWITTER_CONSUMER_SECRET")
TWITTER_ACCESS_TOKEN = Sys.getenv("TWITTER_ACCESS_TOKEN")
TWITTER_ACCESS_TOKEN_SECRET = Sys.getenv("TWITTER_ACCESS_TOKEN_SECRET")
```

Setup the Authorisation for Twitter

```
setup_twitter_oauth(
  TWITTER_CONSUMER_KEY,
  TWITTER_CONSUMER_SECRET,
  TWITTER_ACCESS_TOKEN,
  TWITTER_ACCESS_TOKEN_SECRET
)
```

```
## [1] "Using direct authentication"
```

Now let's find out the tweets regarding "datascience"

```
DStweet <- searchTwitter("BigData",n=100)
```

Convert the list to Data Frame

```
dfTweet <- twListToDF(DStweet)
```

Thesea are the various details we can see regarding the tweets

Let's try to see what we have in the data frame

```
tbl_df(dfTweet[,-1]) # removing the "text" as it skews the display of the output
```

```
## Source: local data frame [100 x 15]
##
##    favorited favoriteCount replyToSN             created truncated
##        (lgl)         (dbl)     (chr)              (time)     (lgl)
## 1      FALSE             0        NA 2016-03-19 09:50:47     FALSE
## 2      FALSE             0        NA 2016-03-19 09:50:47     FALSE
## 3      FALSE             0        NA 2016-03-19 09:50:47     FALSE
## 4      FALSE             0        NA 2016-03-19 09:50:46     FALSE
## 5      FALSE             0        NA 2016-03-19 09:50:45     FALSE
## 6      FALSE             0        NA 2016-03-19 09:50:45     FALSE
## 7      FALSE             0        NA 2016-03-19 09:50:44     FALSE
## 8      FALSE             0        NA 2016-03-19 09:50:43     FALSE
## 9      FALSE             0        NA 2016-03-19 09:50:42     FALSE
## 10     FALSE             0        NA 2016-03-19 09:50:42     FALSE
## ..       ...           ...       ...                 ...       ...
## Variables not shown: replyToSID (lgl), id (chr), replyToUID (chr),
##   statusSource (chr), screenName (chr), retweetCount (dbl), isRetweet
##   (lgl), retweeted (lgl), longitude (lgl), latitude (lgl)
```

We can now do a lot with this info, e.g. find the most re-tweeted text, most active user, which words,tweets are most liked, etc. I will be adding more code to this script. Stay tuned!