

Mining twitter with R



@author Chengjun WANG

@date July 22, 2011

0. Load the package

- `library(twitteR)`
 - `sessionInfo()` # See the information of packages in use.
 - `# update.packages()` # press enter to skip, and press 'y' to choose `twitteR`
 - `# Only twitteR_0.99.9 can run it.`
-

1. twitter search

- `tweets = searchTwitter("#rstats", n=1500)`
 - # The `n=1500` specifies the maximum number of tweets supported by the Search API
 - # `head(tweets)` # return first 6 tweets searched
 - # `class(tweets[[1]])`
 - `length(tweets)`
 - `class(tweets)`
 - `tweet=tweets[[1]]`
 - `name<-tweet$getScreenName()`
 - `name$getLocation()`
 - `tweet$getText()`
-

use plyr

- `library(plyr)`
 - `tweets.df = ldply(tweets, function(t) t$toDataFrame())`
 - `# ldply: split list, apply function, and return results in a data frame.`
 - `# str(tweets.df)`
 - `tweets.text = laply(tweets, function(t) t$getText())`
 - `tweets.name1 = laply(tweets, function(t) t$getScreenName())`
 - `# compare ldply with laply`
 - `tweets.name2 = ldply(tweets, function(t) t$getScreenName())`
 - `head(tweets.name1, 2)`
 - `head(tweets.name2, 10)`
-

2. publicTimeline

- `publicTweets <- publicTimeline()`
 - `length(publicTweets)`
 - `publicTweets[1:5]`
 - `publicTweets[[1]]$getScreenName() # isS4(publicTweets[[1]])`
 - #control the period, place
 - `searchTwitter('charlie sheen', since='2011-03-01', until='2011-07-12', n=100)`
 - `searchTwitter('charlie sheen', since='2011-03-01',n=10)`
 - `searchTwitter('patriots', geocode='42.375,-71.1061111,10mi')`
 - `searchTwitter("#beer", n=100)`
 - `Rtweets(n=37)`
-

3. Trending topic

- `curTrends <- getTrends("current")`
 - `yesterdayTrends <- getTrends("daily", date = as.character(Sys.Date()-1))`
 - `lastWeekTrends <- getTrends("weekly", date = as.character(Sys.Date()-7))`
 - `#length(lastWeekTrends)`
-

4. Get a particular user's timeline

- `crantastic <- getUser("crantastic")`
 - `ChengjunWANG <- getUser("ChengjunWANG")`
 - `cranTweetsLarge <- userTimeline("cranatic", n = 100)`
 - `cjwTweets <- userTimeline("ChengjunWANG", n = 100)`
 - `# Error in .self$twFromJSON(out) : Error: Not authorized`
-

Get the information source of tweets

- `sources <- sapply(publicTweets, function(x) x$getDataSource())`
 - `sources_1 <- gsub("", "", sources)`
 - `sources_2 <- strsplit(sources_1, ">")`
 - `sources_3 <- sapply(sources_2, function(x) ifelse(length(x) > 1, x[2], x[1]))`
 - `pie(table(sources_3))`
-

Authentication with OAuth

- # ONLY R2.13 CAN INSTALL RJSONIO
 - library(ROAuth)
 - cred <- OAuthFactory\$new(consumerKey = 'your key',
 - consumerSecret = 'your secret',
 - requestURL= 'http://api.twitter.com/oauth/request_token',
 - accessURL= 'http://api.twitter.com/oauth/access_token',
 - authURL= 'http://api.twitter.com/oauth/authorize')
 -
 - # cred\$handshake()
 - # The OAuth object, once the handshake is complete, can be saved and reused.
 - # You should not ever have to redo the handshake.
-

Reference

- Jeff Gentry. Twitter client for R. June 14, 2011
 - Jeff Gentry. Package 'twitterR'. June 15, 2011
 - <http://jeffreybreen.wordpress.com/2011/07/21/one-liners-twitter/>
-

The End

- Jeff Gentry has done quite good job in this package named `twitterR`, which make great progress in the edition 0.99.9, however, there seems to be a long way to go.
 - The potential is great, and I will pay more attention to the future improvement of it.
-