

Twitter client for R

Jeff Gentry

November 14, 2011

1 Disclaimer

Because vignettes are built at various points of time (often automatically), and because a lot of the examples are pulling live data from Twitter at the time of being built, it is possible that some of the content in the examples of this document will be unsavory. I've tried to use users and feeds that are unlikely to be this way, but particularly when looking at the public timeline all bets are off.

2 Introduction

Twitter is a popular service that allows users to broadcast short messages ('*tweets*') for others to read. These can be used to communicate with friends, to display headlines, for restaurants to list daily specials, and more. The *twitteR* package is intended to provide access to the Twitter API within R. Users can make access large amounts of Twitter data for data mining and other tasks.

When joined with the *ROauth* package, this package can be used to push the API further and directly interact by posting tweets, dealing with direct messages and enjoying enhanced API rate limitations.

3 Initial Notes

3.1 Support mailing list

While this package doesn't generate a huge volume of emails to me, I have found that the same questions tends to come up repeatedly (often when something has been broken!). I also field requests for advice on practical application of this package which is an area that I'm far from expert at. I've set up a mailing list to better manage emails from users as this way, with the idea being that there'll now be a searchable archive and perhaps other users might be able to chime in. The URL for this mailing list is <http://lists.hexdump.org/listinfo.cgi/twitter-users-hexdump.org>

3.2 Notes for this version

The reworking is still underway. The primary change for this version has been better support for R-2.14

3.3 Notes on API coverage

The ultimate goal is to provide full coverage of the Twitter API, although this is not currently the case. Aspects of the API will be added over time, although if there are particular places that you find missing, please contact me.

3.4 Notes on the classes

There are five classes in this package: `user`, `status`, `trend`, `rateLimitInfo`, and `directMessage`. As of this version they have all been implemented as reference classes (see `setRefClass`). The first two were previously implemented as S4 classes. To help maintain backwards compatibility, the S4 methods (all accessors) have been left in for those two classes although new code should be using the new style accessors.

4 Getting Started

We'll focus first on those sections of the package that do not require `ROAuth` authentication. The rest of this document won't be an encyclopedic report on the functionality of the package but will just show some basic techniques.

```
> library(twitteR)
```

5 Exploring Twitter

A Twitter *timeline* is simply a stream of tweets - this might be the *public timeline* which is comprised of all public tweets, it might be a user's timeline which would be all of their tweets, or it might even be a timeline to look at one's friend's tweets. Just as there are various *timelines* in Twitter, the *twitteR* package provides various interfaces to access them. The first and most obvious would be the *public timeline*, which retrieves the 20 most recent public tweets on Twitter, returned to the user as a list of *status* objects.

```
> publicTweets <- publicTimeline()
> length(publicTweets)
```

```
[1] 20
```

```
> publicTweets[1:5]
```

```

[[1]]
[1] "rosalagos: Ciclo CinePsicoan\xellisis desde Lacan. El orden simb\x3lico en el Siglo XX

[[2]]
[1] "Antoniamulisani: Nhac a vi fez brigadeiro para me deixar feliz :3"

[[3]]
[1] "Nachesjewelry: http://t.co/QU1ayPvh http://t.co/vp8pcuil"

[[4]]
[1] "Jay_NoDaysOff: Nobody really cares if you\031re miserable, so you might as well be happ

[[5]]
[1] "dira_vanessa: Theyre ready! http://t.co/YqGfW9Fz"

> publicTweets[[1]]$getScreenName()

[1] "rosalagos"

> publicTweets[[1]]$getCreated()

[1] "2011-11-15 01:42:01 UTC"

> publicTweets[[1]]$getText()

[1] "Ciclo CinePsicoan\xellisis desde Lacan. El orden simb\x3lico en el Siglo XXI, no es m

Similarly, we can look at a particular user's timeline. This will only work
properly if that user has a public account or you are authenticated and have
access to that account, and can take either a user's name or an object of class
user (more on this later). For this example, let's use the user cranatic.

> cranTweets <- userTimeline('cranatic')
> cranTweets[1:5]

[[1]]
[1] "cranatic: Update: CITAN, coxme, dclone, dynaTree, FAWr, mefa, mefa4, memisc, mosaic, nl

[[2]]
[1] "cranatic: New: bsml. http://bit.ly/k5mVN2 #rstats"

[[3]]
[1] "cranatic: Update: diversitree, gtcorr, raster, RgoogleMaps, tileHMM, tm.plugin.dc, WGCN

[[4]]
[1] "cranatic: Update: DoE.wrapper, Epi, GeneReg, R2Cuba, SubpathwayMiner. http://bit.ly/90f

[[5]]
[1] "cranatic: Update: GenABEL, Mifuns, Mifuns, OrdFacReg, PairViz, RExcelInstaller, SPOT. h

```

By default this command returns the 20 most recent tweet. As with most (but not all) of the functions, it also provides a mechanism to retrieve an arbitrarily large number of tweets up to limits set by the Twitter API, which vary based on the specific type of request. (warning: At least as of now there is no protection from overloading the API rate limit so be reasonable with your requests).

```
> cranTweetsLarge <- userTimeline('cranatic', n=100)
> length(cranTweetsLarge)

[1] 100
```

5.1 Searching Twitter

The `searchTwitter` function can be used to search for tweets that match a desired term. Example searches are such things as hashtags, basic boolean logic such as AND and OR. The `n` argument can be used to specify the number of tweets to return, defaulting to 25.

```
> sea <- searchTwitter('#twitter', n=50)
> sea[1:5]

[[1]]
[1] "mylainz: anyone !!back to basic ang #twitter ??"

[[2]]
[1] "guui_bonfanti: Eae #TWITTER ,"

[[3]]
[1] "SVSalon: RT @fernhilldammit: Hey. Creepy #activity gone. Mentions and retweets back. Ya

[[4]]
[1] "MiAngel_Millan: Me pidieron una donaci\xfa3n economica en #Twitter y me mandan su cuenta

[[5]]
[1] "Ksdude1: #homework sucks and so does #twitter can't check my #mentions"
```

5.2 Looking at users

To take a closer look at a Twitter user (including yourself!), run the command `getUser`. This will only work correctly with users who have their profiles public, or if you're authenticated and granted access.

```
> crantastic <- getUser('crantastic')
> crantastic

[1] "Crantastic"
```

5.3 Trends

Twitter keeps track of topics that are popular at any given point of time, and allows one to extract that data. We're able to see the 20 trending topics per hour for a given day, or the 30 trending topics per day for a given week.

```
> yesterdayTrends <- getTrends('daily', date=as.character(Sys.Date()-1))
> length(yesterdayTrends)
```

```
[1] 480
```

```
> lastWeekTrends <- getTrends('weekly', date=as.character(Sys.Date()-7))
> length(lastWeekTrends)
```

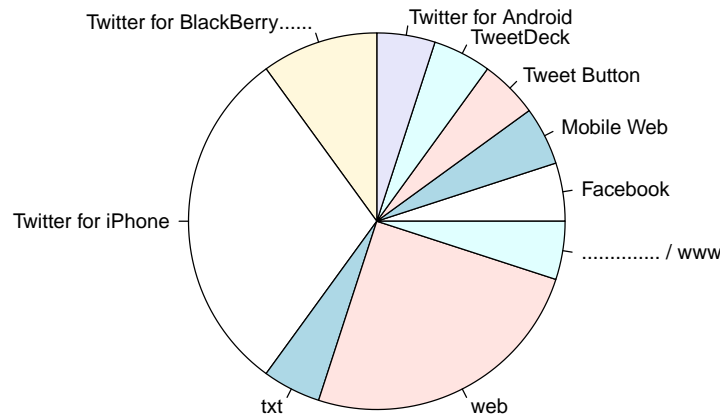
```
[1] 210
```

5.4 A simple example

Just a quick example of how one can interact with actual data. Here we will pull the most recent results from the public timeline and see the clients that were used to post those statuses. We can look at a pie chart to get a sense for the most common clients.

Note that sources which are not the standard web interface will be presented as an anchored URL string (`<A>...`). There are more efficient means to rip out the anchor string than how it is done below, but this is a bit more robust for the purposes of this vignette due to issues with character encoding, locales, etc.

```
> sources <- sapply(publicTweets, function(x) x$statusSource())
> sources <- gsub("</a>", "", sources)
> sources <- strsplit(sources, ">")
> sources <- sapply(sources, function(x) ifelse(length(x) > 1, x[2], x[1]))
> pie(table(sources))
```



5.5 Conversion to data.frames

There are times when it is convenient to display the object lists as an `data.frame` structure. To do this, every class has a reference method `toDataFrame` as well as a corresponding S4 method `as.data.frame` that works in the traditional sense. Converting a single object will typically not be particularly useful by itself but there is a convenience method to convert an entire list, `twListToDF` which takes a list of objects from a single *twitter* class:

```
> df <- twListToDF(publicTweets)
> df[1:3,1:3]
```

```
1 Ciclo CinePsicoan\ellisis desde Lacan. El orden simb\lico en el Siglo XXI, no es m\el
2
3
   favorited replyToSN
1      TRUE        NA
2      TRUE        NA
3      TRUE        NA
```

6 Authentication with OAuth

OAuth is an authentication mechanism gaining popularity which allows applications to provide client functionality to a web service without granting an end user's credentials to the client itself. This causes a few wrinkles for cases like ours, where we're accessing Twitter programmatically. The *ROAuth* package can be used to get around this issue.

The first step is to create a Twitter application for yourself. Go to <https://twitter.com/apps/new>. Set the "Application Type" as "Client", and "Default Access Type" as "Read & Write". This will provide you with two strings, a consumer key and a consumer secret. Record these for your future use.

Three other pieces of information you will need:

- *requestURL*: `https://api.twitter.com/oauth/request_token`
- *accessURL*: `http://api.twitter.com/oauth/access_token`
- *authURL*: `http://api.twitter.com/oauth/authorize`

In your R session, you'll want to do the following:

```
> cred <- OAuthFactory$new(consumerKey=YOURKEY,
+                           consumerSecret=YOURSECRET,
+                           requestURL=requestURL,
+                           accessURL=accessURL,
+                           authURL=authURL)
> cred$handshake()
```

At this point, you'll be prompted with another URL, go to that URL with your browser and you'll be asked to approve the connection for this application. Once you do this, you'll be presented with a PIN, enter that into your R session. Your object is now verified.

Lastly, to use that credential object within an R session, use the `registerTwitterOAuth` function. Passing your `OAuth` object to that function will cause all of the API calls to go through Twitter's OAuth mechanism instead of the standard URLs:

```
> registerTwitterOAuth(cred)
```

The `OAuth` object, once the handshake is complete, can be saved to a file and reused. You should not ever have to redo the handshake unless you remove authorization within the Twitter website.

7 Session Information

The version number of R and packages loaded for generating the vignette were:

```
R version 2.14.0 (2011-10-31)
Platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit)

locale:
[1] C/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods   base

other attached packages:
[1] twitterR_0.99.13 RJSONIO_0.96-0  RCurl_1.6-10    bitops_1.0-4.1

loaded via a namespace (and not attached):
[1] tools_2.14.0
```