Twitter client for R

Jeff Gentry

September 28, 2010

1   IMPORTANT NOTICE REGARDING OAUTH

The old Twitter authentication mechanism is being removed in August 2010 in favor of OAuth. As of this writing there is not an OAuth solution for R, so unless someone (or myself) writes such a beast you will not be able to access any authenticated aspects of Twitter. Many functions that you might be familiar with using will no longer work properly, if at all. These have been set to be defunct.

Also not that it is not simply a matter of having an OAuth interface, but also that Twitter's usage of OAuth is very tied to having stable applications as opposed to scripts being written. There are currently various potential workarounds for this that I'm looking into.

2   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.

3   Introduction

Twitter is a popular service that allows users to broadcast short messages ('*tweets*') for others to read. These can be used to communicate wtih 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. As of this writing, posting tweets, direct messaging and many other tasks are not supported due to lack of authentication options.

4 Getting Started

As of this writing, the following functionality is supported: Searching Twitter, scanning the public timeline, looking at the timeline of a specific user and looking at the followers and followees of a specific user. Due to the lack of authentication, protected accounts and tweets will not be visible to the *twitteR* package.

```
> library(twitteR)
```

5 Time to talk about timelines

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.

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

[1] 20

> publicTweets[1:5]

[[1]]
[1] "ScoreHints: #Sharks do lead at half time. Get the hint? WWW.SCOREHINTS.COM"

[[2]]
[1] "1031dactly: ãÃÑãĄ■LàÿřãĆ■LãĄ■lãĄŢãĄĎãÃĆãĄ■LãĄţãĆ■ãĄńãĄŹãĆŃïij■ãĄ■LãĄţãĆ■ãĄńãĄŹãĆŃïij■ãĄ

[[3]]
[1] "urink: segun comentaristas de ESPN pondran la defensa de salcido desmintiendo lo del tv

[[4]]
[1] "eX_Hooper24: Lmmfao RT @Luv_IsLove: @eX_Hooper24 fuck dat! Dey some damn donuts for it

[[5]]
[1] "HEY2san: http://bit.ly/cEeYnr"
```

Similarly, we can look at a particular user's timeline. This will only work properly if that user has a public 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: DoE.wrapper, Epi, GeneReg, R2Cuba, SubpathwayMiner. http://bit.ly/90f

[[2]]
[1] "cranatic: Update: GenABEL, MIfuns, MIfuns, OrdFacReg, PairViz, RExcelInstaller, SPOT. h

[[3]]
[1] "cranatic: New: CFL. http://bit.ly/braBwy #rstats"

[[4]]
[1] "cranatic: Update: R.matlab, RcmdrPlugin.qual, SMCP. http://bit.ly/9wYOju #rstats"

[[5]]
[1] "cranatic: New: LVQTools. http://bit.ly/9wYOju #rstats"
```

By default this command returns the 20 most recent tweets, as is common with all of these functions. As with most (but not all) of the functions, it also provides a mechanism to retrieve an arbitrarily large number of tweets (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", num = 50)
> sea[1:5]

[[1]]
[1] "SaJiiD_GaGa: RT @Neztor_annie: @SaJiiD_GaGa #TWITTER  #TWITTER  #TWITTER  #TWITTER// ja

[[2]]
[1] "LaNaniSh: Dicen que @comediapolitica sera diputado del PRD y dejara #twitter #diadelbor

[[3]]
[1] "shakenandfaint: RT @BorisLizana: El 99% de #Twitter ya tiene el #NewTwitter, si eres el

[[4]]
[1] "carlosarturo521: @BrendaMolinito  bueno tu que te crees prometiste no descuidar #Twitte
```

```
[[5]]
[1] "jorgerdrgx: Por que me hablo por #twitter con @DavidBste y @camiloz81, si siempre los t
```

5.2 Seeing what other R folks are up to

The `Rtweets` function will retrieve the 25 most recent tweets that carry the
rstats hash tag, which is commonly used by members of the *R* community. As
with other functions, the number of returned tweets can be modified.

```
> rt <- Rtweets(n = 50)
> rt[1:5]

[[1]]
[1] "imusicmash: machine learning books http://metaoptimize.com/qa/questions/186/ #rstats #d

[[2]]
[1] "sean_lee87: For me, much easier 2 learn from worked examples; online help gives me list

[[3]]
[1] "berndweiss: sounds cool: metaSEM conducts univariate/multivariate meta-analyses using a

[[4]]
[1] "jduckles: GEOS wrapper comes to #rstats http://bit.ly/dnwg8X"

[[5]]
[1] "freakonometrics: Detecting distributions with infinite mean, #rstats, http://tinyurl.co
```

6 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.

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

[1] "Crantastic"
```

Furthermore, we can look at this user's friends, as well as those following
them (same disclaimer regarding public profiles applies here):

```
> friends <- userFriends("crantastic")
> friends[[1]]

[1] "IMKristenBell"

> followers <- userFollowers("crantastic")
> followers[1:5]
```

```
[[1]]
[1] "nerdibird"

[[2]]
[1] "sheweeherman"

[[3]]
[1] "thenetnat"

[[4]]
[1] "keithkurson"

[[5]]
[1] "rebecca_a_moses"
```

6.1 The user class

In both of the above cases, the argument can be a string noting the user's screen name, or a *user* object. Let's look more closely at the *user* class to see what is available within it.

The following is a look at the collection of available get methods for the *user* class:

```
> curUser <- friends[[1]]
> screenName(curUser)

[1] "IMKristenBell"

> description(curUser)

[1] "5'1 is the new 6'2"

> tweetCount(curUser)

[1] 646

> followersCount(curUser)

[1] 191820

> favoritesCount(curUser)

numeric(0)

> friendsCount(curUser)

[1] 33

> name(curUser)
```

```
[1] "Kristen Bell "

> protected(curUser)

[1] FALSE

> verified(curUser)

[1] TRUE

> location(curUser)

[1] "Los Angeles, California "

> id(curUser)

[1] 53297035

> lastStatus(curUser)

[1] "Unknown: @jeremy_hopper here u go! Thank u for retweeting! http://tinyurl.com/2v3uepj"
```

7   Dissecting a tweet

The *status* class has the following methods defined:

- *text: Retrieves the text of the tweet*
- screenName: Screen name of the sender
- *id: Retrieves the ID of the tweet*
- created: Retrieves the date the tweet was created, in POSIX date format
- *replyToSN: If this is a reply, the screen name for the reply*
- replyToSID: If this is a reply, the message this is in reply to
- *replyToUID: If this is a reply, the user this is in reply to*
- favorited: If this reply is favorited
- *statusSource: Source of the tweet*

8 Session Information

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

```
R version 2.11.1 (2010-05-31)
x86_64-apple-darwin9.8.0

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

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

other attached packages:
[1] twitteR_0.9.1  RJSONIO_0.3-1  RCurl_1.4-4    bitops_1.0-4.1
```