Twitter en R

Freddy F. Tapia C.

19/02/2020

Extracción de tweets usando R

A continuación se explicará el proceso de como extraer información (tweets) a partir de Twitter mediante dos formas diferentes. La data obtenida puede ser utilizada para diversos propósitos, entre ellos es posible realizar un análisis de sentimientos a la misma con el fin que saber que es lo que piensa las personas o un usuario en específico sobre un tema en particular. En el proceso se usará el paquete "rtweet" de R y un complemento de Google Sheets denominado "Twitter Archiver".

Usando API de Twitter

Para usar la API (Application Programming Interface) de Twitter, es necesario contar con una "app" en Twitter la cual nos proporcionará cuatro valores (claves) que van a ser requeridos por el paquete "rtweet". El proceso para obtener un "app" en Twitter es el siguiente,

1) Contar con una cuenta en Twitter. De no contar con ella, puede crearla en la siguiente dirección.
2) Crear una "app" en Twitter Developers, para ingresar haga click aquí. Luego de esto seleccionar la pestaña "Apps" y rellenar el formulario.
3) Al generarse la app, es necesario crear un token, para ello hay que seleccionar la opción "Create my access token" ("Generar mi token de acceso").
4) Luego de esto se generarán las siguientes claves:

- Consumer key (clave consumidor)
- Consumer secret (secreto consumidor)
- Access token (token de acceso)
- Access token secret (secreto token de acceso)

Una vez obtenida esta información, ya es posible realizar consultas, directamente desde R.

Extracción de tweets de un usuario en específico

La función que nos ayudará a extraer los tweets será "get_timeline", cuyos parámetros principales son,

- **user**: nombre de usuario a consultar.
- **n**: número de tweets a extraer cada vez, el valor por defecto es 100. Este valor no debe exceder los 3200.
- **parse**: valor lógico que indica si el retorno debe ser un data.frame (TRUE) ó una lista (FALSE).
- **check**: valor lógico que indica si se quiere comprobar la tasa límite disponible.

Para lograr esta función logre acceder a la data hay que usar la función "create_token", a la cual se le deben pasar las cuatro claves generadas anteriormente, las cuales son "consumer_key", "consumer_secret", "access_token" y "access_secret", ademas de suministrar el valor del "appname".

```
twitter_token <- create_token(app = appname, consumer_key = key,
                              consumer_secret = secret,
                              access_token = acces_t ,access_secret = access_s)


datos <- get_timeline(user = "@elonmusk", n = 3200, parse = TRUE,
                          check = TRUE)

datos$text[1:5]
```

```
## [1] "@Kitty_McConnell Glad you’re all ok! <U+2764><U+FE0F>"
## [2] "@PicklePunchD @LaurenRow5 @archillect Tim Curry ftw"
## [3] "@archillect <U+0001F525><U+0001F525> makeup"
## [4] "@Smerity Unfortunately, I must agree that these are reasonable concerns"
## [5] "@carkgirl @tesletter True, that was pretty impressive"
```

Un vistazo a la data extraida se presenta a continuación,

```
#PRIMEROS DATOS
head(datos$text)
```

```
## [1] "@Kitty_McConnell Glad you’re all ok! <U+2764><U+FE0F>"
## [2] "@PicklePunchD @LaurenRow5 @archillect Tim Curry ftw"
## [3] "@archillect <U+0001F525><U+0001F525> makeup"
## [4] "@Smerity Unfortunately, I must agree that these are reasonable concerns"
## [5] "@carkgirl @tesletter True, that was pretty impressive"
## [6] "@tesletter My conversations with Gates have been underwhelming tbh"
```

```
#NOMBRES DE COLUMNAS
names(datos)
```

```
##  [1] "user_id"                "status_id"
##  [3] "created_at"             "screen_name"
##  [5] "text"                   "source"
##  [7] "display_text_width"     "reply_to_status_id"
##  [9] "reply_to_user_id"       "reply_to_screen_name"
## [11] "is_quote"               "is_retweet"
## [13] "favorite_count"         "retweet_count"
## [15] "quote_count"            "reply_count"
## [17] "hashtags"               "symbols"
## [19] "urls_url"               "urls_t.co"
## [21] "urls_expanded_url"      "media_url"
## [23] "media_t.co"             "media_expanded_url"
## [25] "media_type"             "ext_media_url"
## [27] "ext_media_t.co"         "ext_media_expanded_url"
## [29] "ext_media_type"         "mentions_user_id"
## [31] "mentions_screen_name"   "lang"
## [33] "quoted_status_id"       "quoted_text"
## [35] "quoted_created_at"      "quoted_source"
## [37] "quoted_favorite_count"  "quoted_retweet_count"
## [39] "quoted_user_id"         "quoted_screen_name"
## [41] "quoted_name"            "quoted_followers_count"
```

```
## [43] "quoted_friends_count"    "quoted_statuses_count"
## [45] "quoted_location"         "quoted_description"
## [47] "quoted_verified"         "retweet_status_id"
## [49] "retweet_text"            "retweet_created_at"
## [51] "retweet_source"          "retweet_favorite_count"
## [53] "retweet_retweet_count"   "retweet_user_id"
## [55] "retweet_screen_name"     "retweet_name"
## [57] "retweet_followers_count" "retweet_friends_count"
## [59] "retweet_statuses_count"  "retweet_location"
## [61] "retweet_description"     "retweet_verified"
## [63] "place_url"               "place_name"
## [65] "place_full_name"         "place_type"
## [67] "country"                 "country_code"
## [69] "geo_coords"              "coords_coords"
## [71] "bbox_coords"             "status_url"
## [73] "name"                    "location"
## [75] "description"             "url"
## [77] "protected"               "followers_count"
## [79] "friends_count"           "listed_count"
## [81] "statuses_count"          "favourites_count"
## [83] "account_created_at"      "verified"
## [85] "profile_url"             "profile_expanded_url"
## [87] "account_lang"            "profile_banner_url"
## [89] "profile_background_url"  "profile_image_url"
```

```
#DIMENSION
dim(datos)
```

```
## [1] 3199   90
```

Extracción de tweets sobre un tema cualquiera

Para la extracción de tweets que contengan una palabra cualquiera, hay que usar la función "search_tweets", cuyos principales parámetros son,

- **q**: query a ser buscado, puede ser una palabra o una cadena de caracteres, donde se especifique las palabras a ser buscadas.
- **n**: número total de tweets a ser extraidos, el valor por defecto es 100. EL valor máximo a extraer por cada token es de 18.000 tweets.
- **type**: caracter que indica reciente ("recent"), mixto ("mixed") ó popular ("popular").

```
#B) BUSCANDO UNA PALABRA CUALQUIERA
apple <- search_tweets(q = "apple",n=100,type = "recent")
```

A continuación se presenta la data extraida,

```
#PRIMEROS DATOS
head(apple$text)
```

```
## [1] "Yuk buat kalian pengguna apple sekarang ga usah bingung mau beli aplikasi atau lagu dimana, kar
## [2] "Unbelievable  https://t.co/IZwPXdH5eB"
```

```
## [3] "Yuk buat kalian pengguna apple sekarang ga usah bingung mau beli aplikasi atau lagu dimana, kar
## [4] "<U+0E08><U+0E30><U+0E23><U+0E49><U+0E2D><U+0E07><U+0E44><U+0E2B><U+0E49><U+0001F602> <U+0E19><U+
## [5] "Organizing isn't just about a candidate. It's about community. @ChrisWestefeld joined us this w
## [6] "https://t.co/as2sa45ZjK"
```

```
#NOMBRES DE COLUMNAS
names(apple)
```

```
##  [1] "user_id"                "status_id"
##  [3] "created_at"             "screen_name"
##  [5] "text"                   "source"
##  [7] "display_text_width"     "reply_to_status_id"
##  [9] "reply_to_user_id"       "reply_to_screen_name"
## [11] "is_quote"               "is_retweet"
## [13] "favorite_count"         "retweet_count"
## [15] "quote_count"            "reply_count"
## [17] "hashtags"               "symbols"
## [19] "urls_url"               "urls_t.co"
## [21] "urls_expanded_url"      "media_url"
## [23] "media_t.co"             "media_expanded_url"
## [25] "media_type"             "ext_media_url"
## [27] "ext_media_t.co"         "ext_media_expanded_url"
## [29] "ext_media_type"         "mentions_user_id"
## [31] "mentions_screen_name"   "lang"
## [33] "quoted_status_id"       "quoted_text"
## [35] "quoted_created_at"      "quoted_source"
## [37] "quoted_favorite_count"  "quoted_retweet_count"
## [39] "quoted_user_id"         "quoted_screen_name"
## [41] "quoted_name"            "quoted_followers_count"
## [43] "quoted_friends_count"   "quoted_statuses_count"
## [45] "quoted_location"        "quoted_description"
## [47] "quoted_verified"        "retweet_status_id"
## [49] "retweet_text"           "retweet_created_at"
## [51] "retweet_source"         "retweet_favorite_count"
## [53] "retweet_retweet_count"  "retweet_user_id"
## [55] "retweet_screen_name"    "retweet_name"
## [57] "retweet_followers_count" "retweet_friends_count"
## [59] "retweet_statuses_count" "retweet_location"
## [61] "retweet_description"    "retweet_verified"
## [63] "place_url"              "place_name"
## [65] "place_full_name"        "place_type"
## [67] "country"                "country_code"
## [69] "geo_coords"             "coords_coords"
## [71] "bbox_coords"            "status_url"
## [73] "name"                   "location"
## [75] "description"            "url"
## [77] "protected"              "followers_count"
## [79] "friends_count"          "listed_count"
## [81] "statuses_count"         "favourites_count"
## [83] "account_created_at"     "verified"
## [85] "profile_url"            "profile_expanded_url"
## [87] "account_lang"           "profile_banner_url"
## [89] "profile_background_url" "profile_image_url"
```

```
#DIMENSION
dim(apple)
```

```
## [1] 100  90
```

Usando Google Sheets

Este método es una gran alternativa en caso que no se cuente con las claves que requiere la API de Twitter. El mismo es un servicio pago de Google Drive, aunque permite realizar una consulta de alguna palabra en específico, de manera gratuita. Para acceder a este complemento y obtener la data que se desea se deben seguir los siguientes pasos,

1) Ingresar a Google Drive y abrir una hoja de cálculo en blanco.
2) Agregar el complemento Twitter Archiver, el cual se encuentra en los complementos de Google Sheets, para ingresar haga click aquí.
3) Una vez agregado el complemento, ir a la pestaña "complementos" de la hoja de cálculo y selecionar "Twitter Archiver", luego de esto elegir "Create Rule".
4) Luego de esto rellenar la información necesaria y dar click en "Create Search Rule".
5) Una vez creada la busqueda, la hoja de cálculo se rellenará de forma automática y la misma podrá ser descargada en diferentes formatos.

Una data de prueba descargada por este método se presenta a continuación,

```
#2) USANDO GOOGLE SHEETS
tweets <- read.csv(paste0(getwd(),"/Datos/tweets.csv"),header = FALSE)
```

```
#PRIMEROS REGISTROS
head(tweets$Tweet_Text)
```

```
## [1] Tweet Text
## [2] Die neuen Apple Zahlen sind da â\200" alle Infos
## [3] Top five MJ song ðŸ\230ŒðŸ\230ŒðŸ\230Œ
## [4]
## [5] Câ\200\231est le meilleur son de son album frÃ¨re
## [6] ** THE #APPLE RULES ** $AAPL Q3 19 Earnings: - Revenue: $53.8B (exp $53.35B) - EPS: $2.18 (exp $1
## 2644 Levels:  ...
```

```
#DIMENSION DEL ARCHIVO
dim(tweets)
```

```
## [1] 3303   21
```

```
#NOMBRES DE COLUMNAS
names(tweets)
```

```
##  [1] "Date"          "Screen names"  "Full name"     "Tweet_Text"
##  [5] "Tweet ID"      "Link(s)"       "Media"         "Location"
##  [9] "Retweets"      "Favorites"     "App"           "Followers"
## [13] "Follows"       "Listed"        "Verfied"       "User Since"
## [17] "Location"      "Bio"           "Website"       "Timezone"
## [21] "Profile Image"
```

Es importante tener en cuenta la principal diferencia que existe en los tres métodos, la misma se centra en la cantidad de tweets extraidos. Por una parte, si se usa la API y se quiere encontrar información de un usuario en específico, 3.200 son la maxima cantidad de tweets que se podrán extraer al realizar una busqueda, aunque es posible elevar este número si se realiza más de una busqueda. Por otra parte, si se usa la API para buscar una palabra o palabras en específico, 18.000 tweets es lo que podremos conseguir, lo cual ya es un número considerable.

Finalmente si se utiiza el complemento de Google Sheets es posible encontrar con una busqueda aproximadamente 3.000 tweets si la misma busqueda se repite varias veces es posible llegar a tener una base de datos con 55.000 tweets. Como se podrá ver cada método tiene sus ventajas y desventajas, así que hay que elegir uno que se adapte a nuestras necesidades.