

# GUIA DE USO: TweetExtract

---

Este documento sirve como guía de instalación y uso de TweetExtract.

## Prerrequisitos

Esta aplicación precisa de un equipo (ordenador de sobremesa o portátil) con conexión a Internet. No es necesario que tenga una gran potencia de cálculo, si bien ésta puede ser altamente recomendable, ya que las tareas de procesamiento de los datos pueden prolongarse en equipos con bajas prestaciones.

Así mismo, se precisa de dos herramientas software, RStudio y R, de una cuenta en la red social Twitter y de una aplicación creada en Twitter Application Management.

## Novedades y limitaciones de TweetExtract

### Principales novedades

- La principal ventaja de TweetExtract es su capacidad de obtención de información de Twitter, su transformación, y posibilidad de descarga y exportación de datos para posteriores análisis.
- Es una aplicación fácil y muy intuitiva, ya que únicamente ejecuta trozos de código de forma secuencial.
- Provee información esencial para un estudio básico de las publicaciones de Twitter, y permite la modificación de diversos parámetros para que el usuario realice su análisis y obtenga los datos de la forma que mejor le convenga.

### Limitaciones

- El principal inconveniente reside en el hecho de que el usuario debe disponer de una cuenta en la red social Twitter para poder realizar una aplicación en “Twitter Application Management”, de tal forma que pueda obtener una serie de tokens (ver más adelante) para poder obtener correctamente los datos de Twitter.
- El usuario precisa de instalar las herramientas software RStudio y R.

## INSTALACIÓN Y USO DE TweetExtract

En este punto se recoge la primera toma de contacto con la herramienta, en la que se enumeran una serie de pasos a seguir y las características que el usuario puede encontrarse en el momento de utilización de esta herramienta. Los pasos para la instalación de TweetExtract incluyen:

1. **[Acceso a la página de Twitter y creación de una cuenta de usuario](https://twitter.com/?lang=es)**

## 2. Creación de una aplicación en [Twitter Application Management](#)

El usuario debe acceder a [Twitter Application Management](#) con los datos de la cuenta de Twitter creada anteriormente (si tenía una cuenta previa este proceso es igualmente válido).

Una vez accede con su cuenta de Twitter, debe proceder a crear una aplicación, para lo cual debe seleccionar “Create New App”.



Ilustración 1 - Creación de Aplicación Twitter

Seguidamente, debe rellenar el formulario que se le mostrará con el nombre de “Application Details” (nombre de la aplicación, descripción de la aplicación y una URL ), aceptar el acuerdo de desarrollo y pulsar “Create your Twitter application”.

### Create an application

**Application Details**

**Name \***  
GuiaUsoAppPrueba  
Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

**Description \***  
Esta es una aplicación de prueba  
Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

**Website \***  
http://appPrueba.es  
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

**Callback URL**  
Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth\_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

**Developer Agreement**  
 Yes, I have read and agree to the [Twitter Developer Agreement](#).

Create your Twitter application

Ilustración 2 - Formulario de Aplicación Twitter

## 3. Obtención de Tokens de la aplicación

Una vez que se crea la aplicación, se accede a un menú completo para poder gestionar las diferentes características de la cuenta. Para poder utilizar la aplicación en R, es

necesario obtener cuatro tokens para la aplicación creada en la página “Twitter Application Management”. Para ello, el usuario debe acceder a la pestaña “Keys and Access Tokens”.

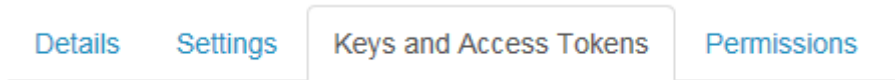


Ilustración 3 - Pestaña de obtención de Tokens

“Keys and Access Tokens” muestra los ajustes de la aplicación. En las dos primeras filas se encontrarán los dos primeros valores que el usuario necesita: “Consumer Key (API Key)” y “Consumer Secret (API Secret)” (Rodeados en la Ilustración 4 - Claves de Aplicación “Consumer Key” y “Consumer Secret” en rojo).



## Application Settings

*Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.*

Consumer Key (API Key)

Consumer Secret (API Secret)

Ilustración 4 - Claves de Aplicación "Consumer Key" y "Consumer Secret"

Así mismo, el usuario debe acceder en la parte inferior de esa misma pantalla a los otros dos valores necesarios, seleccionando para ello “Create my Access token”.

## Your Access Token

*You haven't authorized this application for your own account yet.*

*By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.*

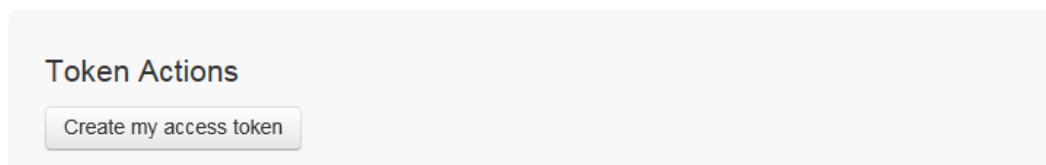


Ilustración 5 - Creación de Tokens de Acceso

Una vez que se generan adecuadamente los tokens, se muestran los dos valores restantes necesarios para el funcionamiento de TweetExtract: “Access Token” y “Access Token Secret”.

## Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token   
Access Token Secret 

Ilustración 6 - Tokens de Acceso

### 4. Descarga de los Datasets

Una vez que el usuario ha realizado los pasos anteriores, deberá abrir RStudio (previamente se debe descargar el IDE [RStudio](#) y el entorno y lenguaje de programación [R](#)).

Para abrir la aplicación el usuario tiene que seleccionar en File>Open File y deberá seleccionar "TweetExtract" en el directorio donde tenga alojada la aplicación. Es importante saber que si el usuario no configura RStudio, por defecto las descargas de los datasets que se podrán realizar se guardarán en el directorio donde se encuentre instalado R. Para que el usuario pueda elegir el directorio donde guardar las descargas de los datasets deberá incluir la ruta en la línea de código correspondiente en la aplicación como se ve en la figura siguiente.

```
#Guardamos en csv en el workspace donde se encuentra instalado R
write.csv(tablaDataset1, file = "datasetTwitter1.csv")

#Guardamos en csv en el directorio que desee el usuario (cambiar directorio1, directorio2 por el directorio que se desee)
write.csv(tablaDataset1, file = "C:/directorio1/directorio2/datasetTwitter1.csv")
```

Ilustración 7 - Directorios de trabajo

Cuando se abre la aplicación en RStudio, el usuario se encuentra el código en la parte superior izquierda, que deberá ir ejecutando paso a paso como se verá más adelante.

Antes de cualquier ejecución del código, el usuario tiene que "conectar" esta aplicación con la aplicación de Twitter que se creó anteriormente y mediante la cual se generaron las cuatro claves/tokens a introducir. Los valores de estas claves se corresponden con los que el usuario generó en el punto 3, y que deberá copiarlos aquí y ponerlos entre comillas (""). Estas claves/tokens se deben introducir en las variables: *consumer\_key*, *consumer\_secret*, *access\_token* y *access\_secret*. A continuación se muestra dónde se deben introducir las claves/tokens.

```
#A continuación introducir las claves y tokens de la aplicación de Twitter
#Sustituir xxxxxxxxxxxxxxx por la clave en cuestión
#Claves
consumer_key <- "xxxxxxxxxxxxxxxxxxxxxxxxxxxx"
consumer_secret <- "xxxxxxxxxxxxxxxxxxxxxxxxxxxx"
access_token <- "xxxxxxxxxxxxxxxxxxxxxxxxxxxx"
access_secret <- "xxxxxxxxxxxxxxxxxxxxxxxxxxxx"
```

Ilustración 8 - Claves/Tokens de Twitter en TweetExtract

Una vez ha realizado la introducción de las claves y tokens, se deberá ejecutar paso a paso el código.

Lo primero que el usuario debe ejecutar son las primeras líneas, encargadas de instalar los paquetes necesarios y preparar el entorno con las librerías adecuadas. Para

ejecutar las líneas de código de la aplicación se debe seleccionar la línea o líneas a ejecutar, y pulsar “CTRL+ENTER” o bien el icono “Run” situado en la parte superior derecha de la ventana donde se encuentra el código.

```
#Aplicacion para la extracción de tweets y su ordenación en una tabla
#Necesitaremos la instalación de estos dos paquetes
install.packages('twitter')
install.packages('ROAuth')

#Importamos las librerías
library(twitter)
library(ROAuth)
```

#### Ilustración 9 - Instalación de paquetes y librerías

Para “conectar” TweetExtract con la aplicación de Twitter se deben ejecutar las líneas de las claves introducidas anteriormente (que se guardaron como variables), así como la propia línea de autenticación.

```
#Twitter authentication
#Es aconsejable pulsar "1" para que se almacenen las claves en el directorio de trabajo
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

#### Ilustración 10 - Autenticación de la aplicación con Twitter

Cuando se ejecutan estas líneas de código aparece en la parte inferior de RStudio una pregunta acerca de si se desea que las credenciales de creación de sesión se almacenen en la carpeta local, cosa que se recomienda debido a la rapidez de ejecución de la aplicación en sucesivas ejecuciones; por tanto, la opción recomendada es “1”.

Seguidamente, se encuentran las líneas de descarga de los timelines de las cuentas de Twitter. El usuario puede descargar tantos timelines como desee, con la salvedad de que para el correcto funcionamiento de la aplicación, éstos deben hacerse de 30 en 30 (debido a las capacidades de procesamiento de RStudio, y a que Twitter no permite realizar un número elevado de acceso a información de cuentas a la vez). La estructura de código encargada de la descarga de los timelines es:

```
tweets1 <- userTimeline("iff_csic", n=50)
tweets2 <- userTimeline("IFICorpuscular", n=50)
tweets3 <- userTimeline("BibliotecaImse", n=50)
tweets4 <- userTimeline("iaaucc", n=50)
tweets5 <- userTimeline("IFISC_mallorca", n=50)
```

#### Ilustración 11 - Descarga de Timelines

Entre comillas y en verde se encuentran los nombres de las cuentas de Twitter que se desea analizar (es importante que el nombre coincida con el nombre de la cuenta en Twitter, aunque no hace falta tener en cuenta mayúsculas y minúsculas), y el valor de “n” indica el número de entradas del timeline que se van a analizar.

Este último dato es importante, ya que como máximo este valor puede ser 3200, pero para el caso del presente proyecto es suficiente con 50, debido a que el análisis que se lleva a cabo con esta herramienta es semanal, y no hay ninguna cuenta con más de 50

entradas en su timeline, con lo que se reducen enormemente los tiempos de procesado de los datos.

Así mismo, es importante tener en cuenta que el valor de datos que se obtiene de los timelines de las cuentas que se analizan incluye retuits a otras cuentas (que posteriormente se eliminarán) así como las propias publicaciones de la cuenta que se analiza.

Como se ha mencionado anteriormente, se pueden analizar hasta 30 cuentas a la vez, por lo que para su ejecución basta con seleccionar las 30 líneas de código (en este caso desde “tweets1” hasta “tweets30”) y se pulsa “CTRL+ENTER” y comenzará su ejecución. Si se deseara analizar más de 30 cuentas una vez que se ejecuten las líneas de “tweets1” a “tweets30” se seguirá con “tweets31” a “tweets60” y así sucesivamente.

En este punto, el usuario tiene los tweets que desea descargados, y la información de los mismos se muestra en la ventana superior derecha como se puede ver en la imagen.

▶ tweets1	List of 28
▶ tweets10	List of 50
▶ tweets11	List of 34
▶ tweets12	List of 19
▶ tweets13	List of 17
▶ tweets135	List of 40
▶ tweets14	List of 26
▶ tweets15	List of 24
▶ tweets16	List of 42
▶ tweets17	List of 17

Ilustración 12 - Lista de Timelines descargados

Lo que se realizará a continuación es una ordenación de los datos que tiene para que se puedan visualizar y descargar en el formato que se desea. Para ello, dispone de 6 líneas de código que se corresponderán a los tweets de “tweets1” a “tweets30”, y que se encargan de:

La primera línea se encarga de unir todos los timelines descargados del 1 al 30 en una sola tabla.

```
#-----Tablas de 1 a 30-----  
#Unión de las tablas en una sola  
tablaTweets1 <-rbind(tweets1,tweets2,tweets3,tweets4...
```

Ilustración 13 - Unión de tablas descargadas

A continuación, se transforma tabla en “bruto” anterior en un *data frame* con el que podemos realizar operaciones y clasificaciones de la información. La línea (df.undup ...) elimina los tweets repetidos.

```
#Convertir tweets a data frame (es una tabla)  
tablaTweets1.df <- twListToDF(tablaTweets1)  
df.undup <- tablaTweets1.df[duplicated(tablaTweets1.df)==FALSE,]
```

Ilustración 14 - Transformación de dataframe de Tweets

La siguiente línea ordena los tweets y las cuentas de los mismos, especificando la fecha de creación del tweet, el número de favoritos y de retuits, el autor, o el texto que contiene el tweet en cuestión.

```
#Especificamos como queremos que sea la tabla del dataframe (los campos que va a tener)
tablaDataset1<-df.undup[,c("id","created","screenName","favoriteCount","retweetCount","
```

Ilustración 15 - Ordenación de Tweets

Finalmente, las dos últimas líneas se centran en visualizar los resultados de los tweets ordenados en una tabla en RStudio, y su descarga en un archivo con formato (.csv).

```
#visualizamos la tabla
view(tablaDataset2)
|
#guardamos en csv
write.csv(tablaDataset2, file = "datasetTwitter2.csv")
```

Ilustración 16 - Visualización y descarga de Tweets

Este proceso se repite sucesivamente para el resto de descargas de timelines, y lo único que debe realizar el usuario es modificar los nombres de “tweets1” a “tweets30” por el nombre de los timelines que le haya dado, así como los nombres de las tablas que se van a ir mencionando como “tablaTweets1” por “tablaTweetsX” siendo “X” el número que desee.

Así mismo, y como se comentó al inicio de este apartado, el usuario puede descargar el dataset generado en formato (.csv), y puede además darle el nombre que desee (en este ejemplo se le ha denominado “datasetTwitter2.csv”) y alojarlo en el directorio que crea conveniente añadiendo el “path” o ruta delante del nombre del dataset (por ejemplo: file = “C:/directorio1/directorio2/nombreDataset.csv”) como se indicó en Ilustración 7 - Directorios de trabajo.

## Resultados

TweetExtract devuelve dos tipos de resultados: los obtenidos en el propio RStudio con el que se realiza el análisis, y aquellos que se descargan tras las transformaciones realizadas (archivo .csv).

Lo que se encuentra el usuario en RStudio (una vez que se ejecuta el comando de visualización de la tabla final) es un ejemplo o forma previa a lo que va a poder descargarse en .csv.

id	created	screenName	favoriteCount	retweetCount	isRetweet	statusSource	text
1	2017-09-12 08:59:59	iff_csic	3	2	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	#Electron Transfer #Helium #droplets @CSICdivulga #...
2	2017-09-13 08:00:41	IFICorpuscular	1	0	FALSE	<a href="http://www.hootsuite.com" rel="nofollow">H...	Esta mañana defiende su #tesis @victorguadilla, sobr...
3	2017-08-28 07:08:34	BibliotecaImse	0	0	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	Model-based implementation of self-configurable inte...
4	2017-09-06 10:22:46	iaaucc	11	13	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	El disco solar (región oculta central) y una eyección de...
5	2017-09-08 12:14:57	IFISC_mallorca	1	0	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	@pushsink @bgoncalves We usually live stream our e...
6	2017-07-12 07:22:52	ITEFL_CSIC	0	0	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	Tecnología de Hipertermia como Terapia en Problemas...
7	2017-09-13 08:56:03	DifusionIFCA	1	1	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	¿Quieres orientar tu carrera profesional a un área llen...
8	2017-09-13 08:06:41	iSpaceSci	0	1	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	We started! @_JFAE https://t.co/SA3QgYCTxj
9	2017-09-11 09:52:18	ebdonana	2	1	FALSE	<a href="http://twitter.com" rel="nofollow">Twitter We...	#EBDpub Parasite biodiversity faces extinction. Los pa...
10	2017-09-13 08:50:07	CREAF_ecologia	0	1	FALSE	<a href="http://www.hootsuite.com" rel="nofollow">H...	Cuando el fósforo pasa de nutriente a contaminante e...

Ilustración 17 - Tabla de final de análisis RStudio

Así mismo, el usuario tiene la posibilidad de descargar estos mismos datos en formato .csv ejecutando la línea final del código. Es importante saber que, una vez descargado el archivo .csv, el usuario se encuentra con los datos almacenados en una única columna y separados todos ellos mediante “comas”, con los que se puede trabajar independientemente de la aplicación y efectuar todos los cambios que considere oportunos.

Esta herramienta admite modificaciones en la forma de descargar tweets y en el procesamiento de los mismos en RStudio, por lo que se puede trabajar con otros paquetes de procesado de la información en cualquier parte de la aplicación, como es el caso de las herramientas de visualización (ggplot2) o de procesamiento de palabras para posibles análisis de sentimiento o de identificación de “hashtags”.