

Máster Universitario en Ciencia de Datos e  
Ingeniería de Computadores:  
**Sistemas de Recuperación de Información y  
de Recomendación.**  
**Análisis en Twitter: Brexit.**

**Autores:**

Francisco Pérez Hernández.  
José Rubén Sánchez Iruela.  
Cristina Zuheros Montes.

— —  
**Profesor:**

Juan Manuel Fernández Luna.

26 de abril de 2017.

# Índice

1. Introducción.	3
2. Conexión a la API de Twitter.	3
3. Extracción y limpieza de Tweets.	4
4. Análisis de sentimientos.	4
5. Clasificación de Tweets.	7
6. Usando Shiny.	9
7. Trending Topics.	9
8. Comparativa iPhone, iPad y Android.	11
9. Resultados/Conclusiones.	12
10. Bibliografía.	13

# Índice de figuras

1. Clasificación de tweets por sentimientos. . . . .	5
2. Clasificación de tweets por polaridad. . . . .	5
3. WordCloud. . . . .	6
4. Tabla del sentimiento positivo/negativo. . . . .	7
5. Histograma del sentimiento positivo/negativo. . . . .	8
6. Gráfico chart. . . . .	8
7. Tweets por usuario. . . . .	8
8. Trending Topics. . . . .	10
9. Comparativa para tweets con/sin imagenes/links. . . . .	11
10. Comparativa por palabras. . . . .	12

# 1. Introducción.

En este proyecto recuperaremos información a partir de la red social Twitter sobre el tema Brexit. Posteriormente, se ha ampliado para poder extraer información sobre cualquier hashtag.

Para ello vamos a ir viendo en las siguientes secciones los distintos procedimientos y técnicas que hemos aplicado. Hemos usado el lenguaje R para la extracción y procesamiento de Tweets y HTML para la maquetación de la aplicación shiny.

Las imágenes que vamos a ir mostrando a lo largo del proyecto corresponden al análisis de 5.000 tweets que se han extraído en el día 26/04/2017.

# 2. Conexión a la API de Twitter.

Para poder utilizar la API de Twitter necesitamos previamente tener una cuenta en esta red social. A continuación tenemos que registrar una aplicación, la cuál nos dará estas credenciales y tokens:

- Consumer Key.
- Consumer Secret.
- Access Token.
- Access Secret.

Con esto ya podemos empezar a extraer tweets a través de nuestra aplicación.

### 3. Extracción y limpieza de Tweets.

Para la extracción y limpieza de tweets hemos utilizado la librería `Twitter`. Para ello, primero hemos tenido que cargar nuestras credenciales para así obtener acceso a la API. De esta forma, con la función `searchTwitter` podemos extraer el número de tweets que deseemos en el idioma seleccionado. Nosotros hemos elegido inglés porque pensamos que es más interesante para el tema que vamos a tratar, el Brexit.

Para realizar la limpieza de tweets, lo primero que hacemos es extraer el texto de dichos tweets y posteriormente aplicamos la función `gsub`. Los pasos de limpieza que realizamos son los siguientes:

- Eliminación de usuarios.
- Eliminación de puntuación.
- Eliminación de los números.
- Eliminación de enlaces web.
- Eliminación de los espacios innecesarios.
- Pasar el texto a minúscula.
- Eliminación de valores perdidos.

### 4. Análisis de sentimientos.

En esta sección vamos a analizar las emociones que se muestran en cada tweet. Utilizaremos las funciones `classify_emotion` y `classify_polarity` del paquete `senti` que nos muestra las emociones del tweet (sorpresa, miedo, tristeza...) y la polaridad de los mismos (positivos, negativos o neutros).

En la imagen 1 podemos ver que hay muchos tweets que presentan el sentimiento `unknown`. Esto se debe a que muchas palabras no consiguen ser bien clasificadas ya que no muestran un sentimiento claro.

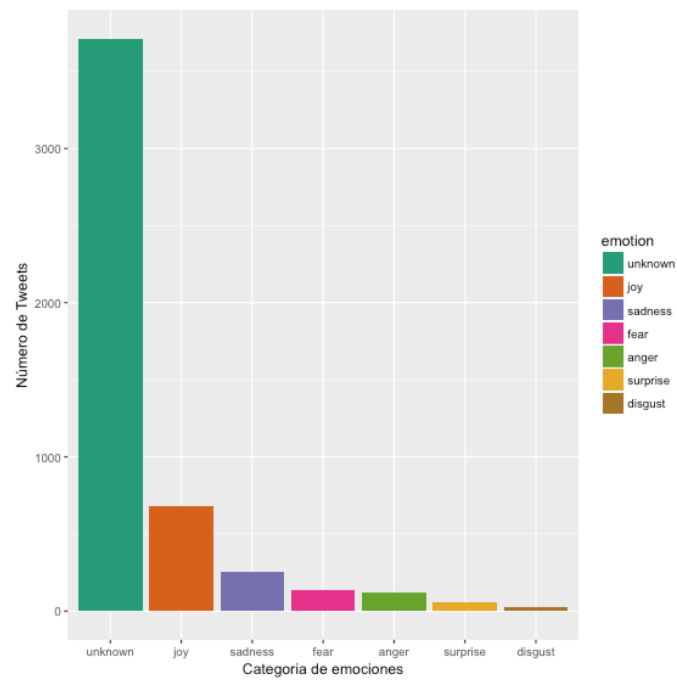


Figura 1: Clasificación de tweets por sentimientos.

En la imagen 2 podemos visualizar la polaridad de los tweets:

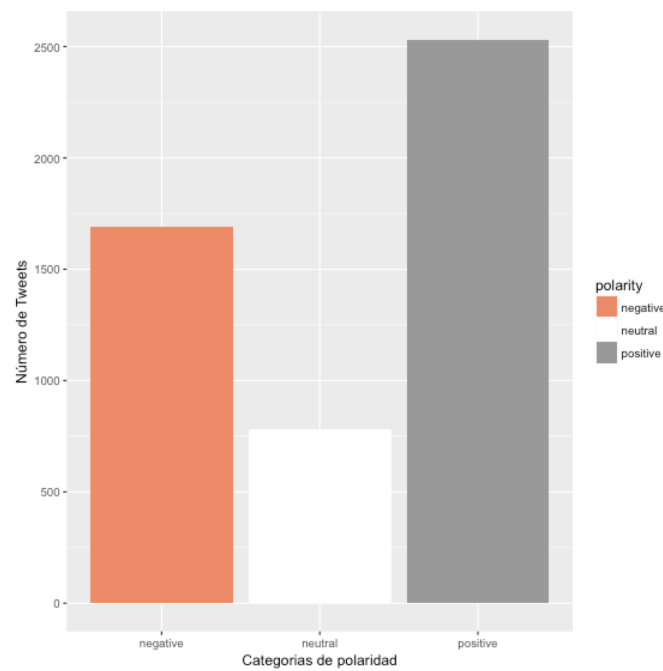
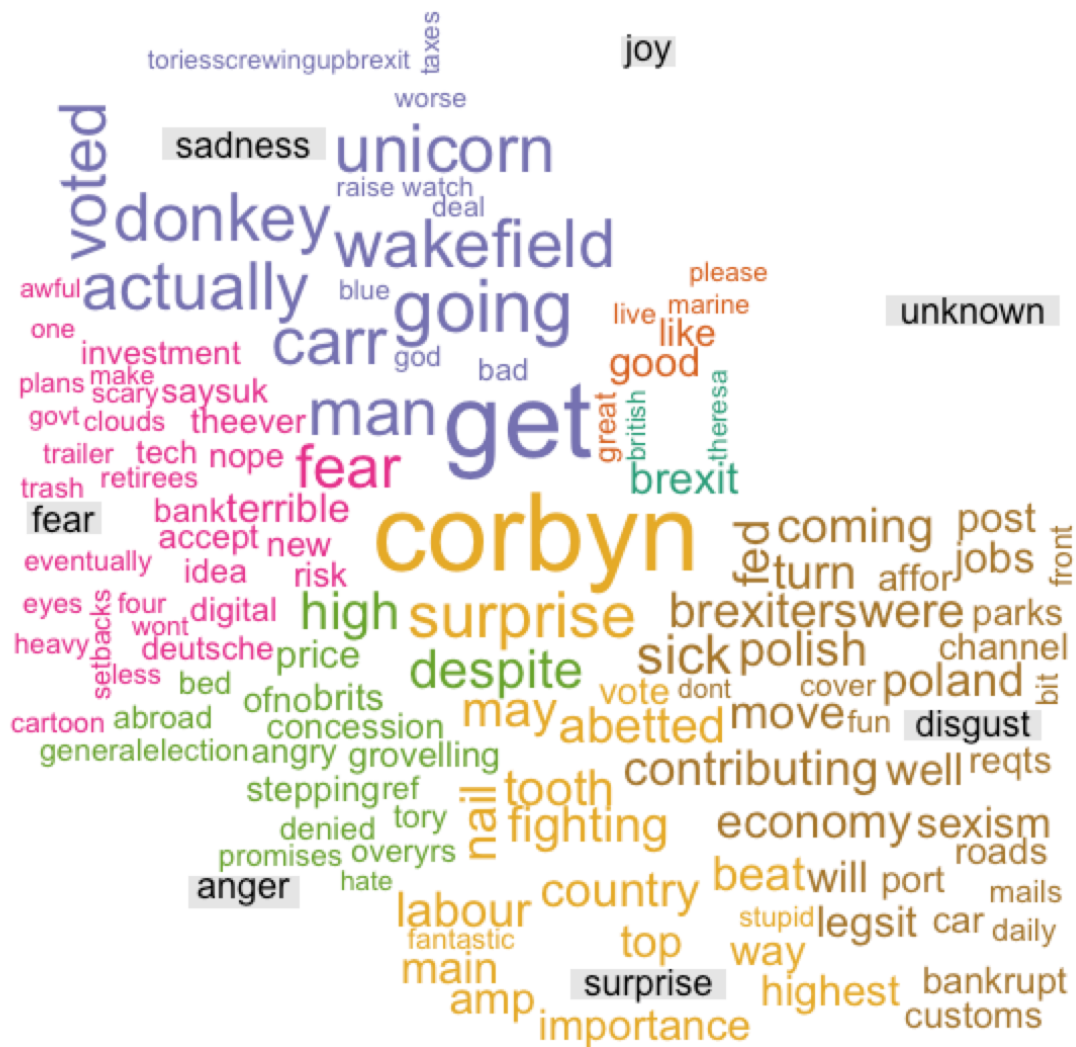


Figura 2: Clasificación de tweets por polaridad.

A partir de las emociones obtenidas y de las palabras más relevantes, vamos a obtener las palabras que más se repiten en cada emoción para crear una nube de palabras, wordcloud. El ejemplo de la nube de palabras se puede ver en la imagen 3:



## 5. Clasificación de Tweets.

Para estudiar la polaridad de los tweets hemos aplicado otra técnica basada en la utilización de diccionarios.

Nos hemos descargado un diccionario de palabras positivas y otro de palabras negativas a los cuáles le hemos añadido algunas palabras más que nos han parecido interesantes. Una vez que tenemos cargados estos diccionarios las utilizamos para puntuar las frases como negativas o positivas, en una escala de cinco, según las palabras que contienen los tweets ya limpiados. Estos diccionarios se pueden conseguir en el siguiente enlace:

Enlace a diccionarios.

A continuación mostramos una tabla de tweets (imagen 4) y los histogramas (imagen 5) que conseguimos obtener con esta técnica. Además, en la imagen 6 podemos ver una clasificación global de los tweets positivos y negativos. En la figura 7 podemos ver los tweets por usuario.

### Representar el sentimiento en forma tabular en una escala de 5

Text	Positive	Negative	Score	PosPercent	NegPercent
RT @acgrayling: No surprise. Corbyn is about Corbyn not country. Labour should be fighting #Brexit tooth & nail: Corbyn has abetted May at	0	3	-3	0.00	100.00
#Brexit Research collaboration and role in post-Brexit free trade agreements vital for universities /aJZ6gvC67W #VoteLeave	1	0	1	100.00	0.00
RT @UKIPNFKN: Anti-fraud body to proceed with investigation into #UKIP's Gill #Brexit #GE2017 /EbnFk8Sd7T	0	1	-1	0.00	100.00
#Brexit Sir Keir Starmer's speech leaves a number of unanswered questions about Labour's Brexit policy /9TX4iuoiMx #VoteLeave	0	0	0	0.00	0.00
RT @moronwatch: "I'm voting #Brexit to stop Poles taking British jobs" - Nestl factory moves to Poland - "Oh crap" /8yV2iPSmpy	0	1	-1	0.00	100.00
'Come and join the #Brexit club' #Farage SHOUTS DOWN Brussels over treatment of eurosceptics /IGJUEeOJFI	0	0	0	0.00	0.00
German officials claim Brexit might be delayed because tiny regional parliaments /vwwKpTww6q #Brexit	0	1	-1	0.00	100.00
Barclays boss sounds Brexit talent warning - BBC News /v1hlEThr4e #Brexit	1	1	0	50.00	50.00
#Brexit causes potholes. Not as impressive as Brexit causes ebola, but desperate times and all that. #remoaners /WclzfK6mQD	1	2	-1	33.33	66.67

Figura 4: Tabla del sentimiento positivo/negativo.

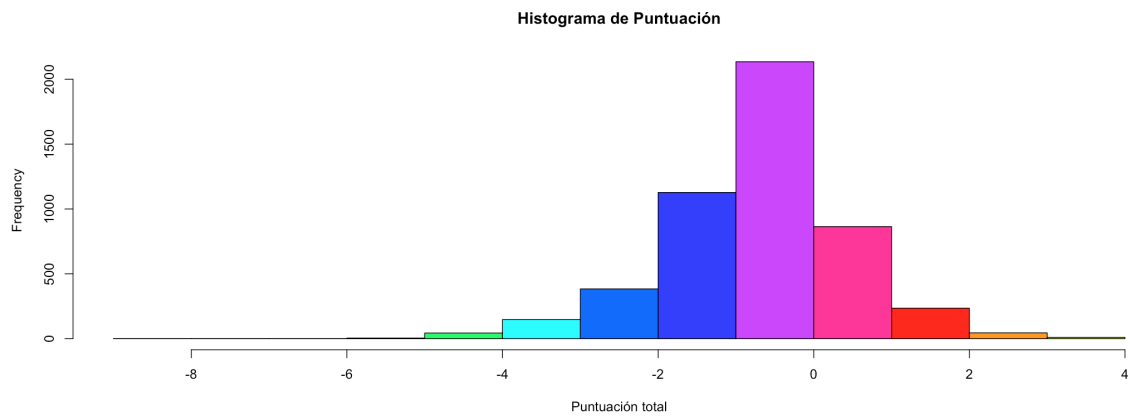


Figura 5: Histograma del sentimiento positivo/negativo.

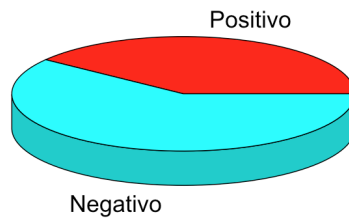


Figura 6: Gráfico chart.

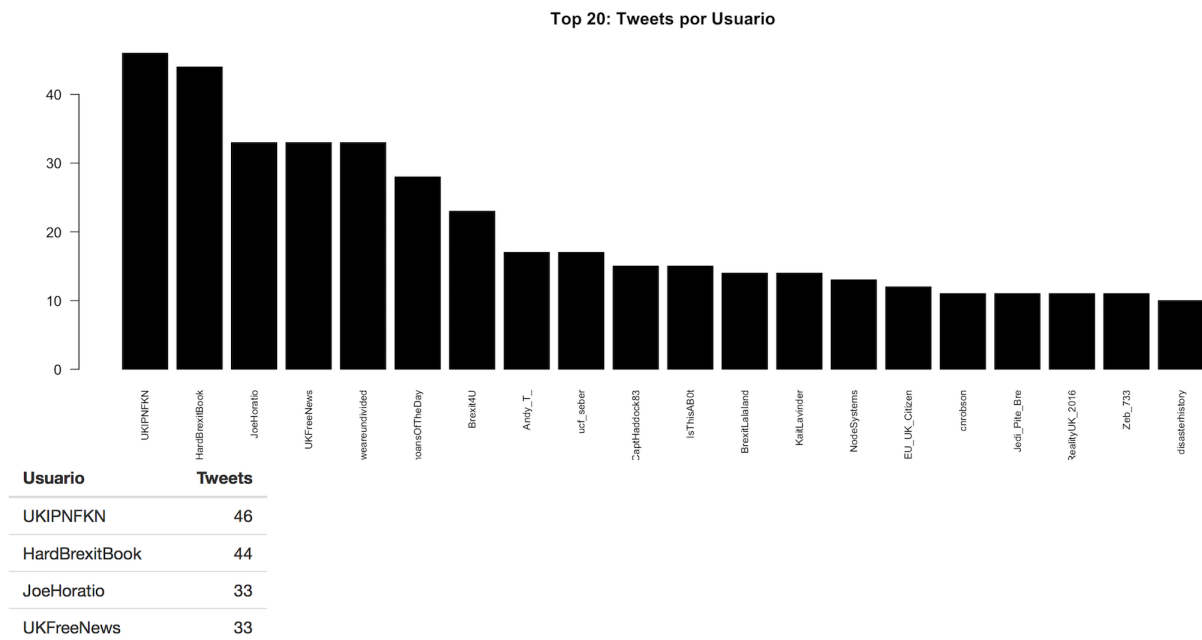


Figura 7: Tweets por usuario.



## 6. Usando Shiny.

Shiny es una plataforma que nos permite compartir aplicaciones web desarrolladas con RStudio. Para ello, necesitamos instalar un conjunto de paquetes en nuestro entorno e instalar nuestro token para así poder subir directamente la aplicación a nuestra cuenta. Este proyecto contará con dos ficheros.

- **Script server.R:** contendrá todo el código en R necesario para el funcionamiento de la aplicación, lo que denominamos el back-end en una aplicación web.
- **Script ui.R:** contendrá el código que permite el desarrollo de la interfaz de nuestra aplicación, en el cuál va mezclado código en R y en HTML, lo que se denomina el front-end.

Hemos usado un servidor externo para poder tener nuestra aplicación Shiny en la web. [En este enlace podemos acceder a la aplicación Shiny.](#)

## 7. Trending Topics.

Otra de las técnicas utilizadas ha sido obtener los trending topics de cualquier parte del mundo y mostrar un listado. Esto se realiza a través de la función `availableTrendLocations` del paquete `TwitterR`. Se puede ver un ejemplo en la imagen 8. Además, en esta imagen ya podemos ver el entorno de trabajo que nos ofrece Shiny.

# Analisis sentimientos Twitter - DATCOM 2017

Introduce el Hashtag por el cual quiere realizar la busqueda '#'

Número de tweets utilizados para el análisis

5 100 1,000

5 105 205 305 405 505 605 705 805 905 1,000

Analizar

Top Trending Tweets de hoy

WordCloud Histogramas Pie Chart

Tabla de tweets Top Tweepers Emociones Polaridad

Comparativa Comparativa 2 Top Hashtags de usuarios

Elige la localización para extraer los trending tweets

Worldwide

Buscar

La siguiente tabla muestra los hashtags de tendencias en Twitter de la ubicación que ha elegido. Estos son los temas candentes de hoy!

data
Jonathan Demme
ESPN
#AlienDay
#whirlpool
#26Abr
Joey Barton
Marshawn Lynch
Echo Look
Vale Roth
#wednesdaywisdom
#FelizMiercoles

Figura 8: Trending Topics.

## 8. Comparativa iPhone, iPad y Android.

Por último realizamos una comparativa sobre los tweets que han sido emitidos a través de dispositivos iPhone, iPad y Android.

Para ello nos hemos fijamos en el campo source de cada tweet que nos permite saber de dónde ha sido emitido. Nos quedamos con estos tres tipos de dispositivos ya que son los más relevantes.

La comparativa se basa en un estudio de los tweets que contienen imágenes o links además de un estudio de las palabras más relevantes que se encuentran en los tweets para cada tipo de dispositivo. Podemos ver algunas de las gráficas en las siguientes 9 y 10.

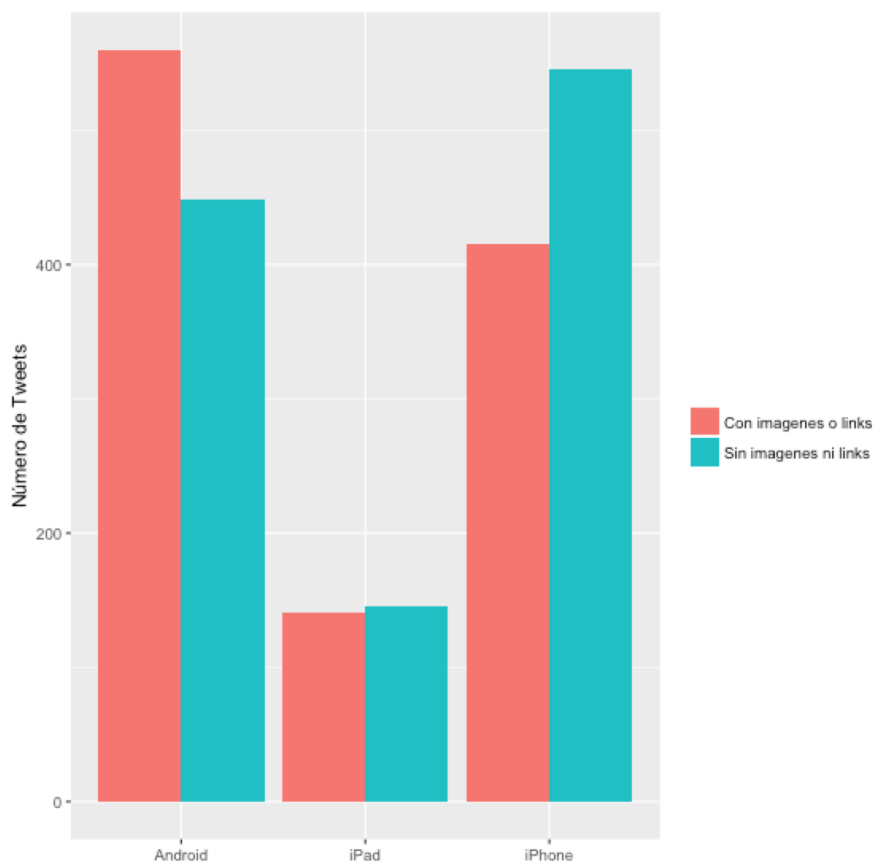


Figura 9: Comparativa para tweets con/sin imagenes/links.

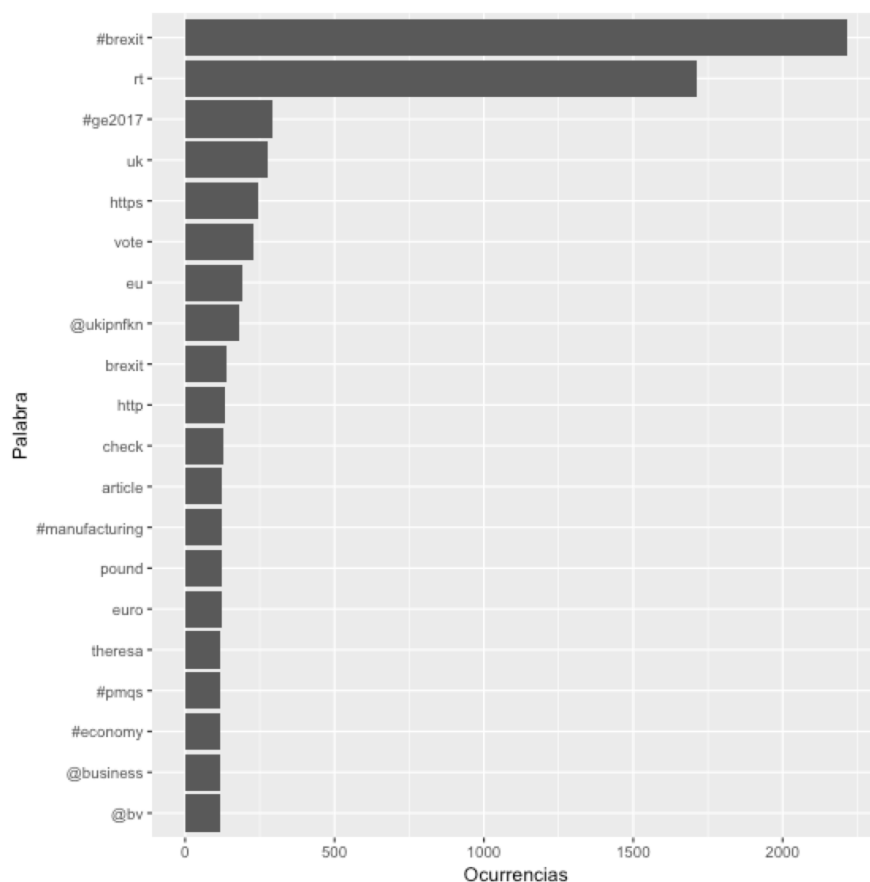


Figura 10: Comparativa por palabras.

## 9. Resultados/Conclusiones.

En este trabajo hemos podido realizar un análisis de sentimientos sobre un hashtag deseado a partir de varias técnicas como es la clasificación de sentimientos o polaridad o mediante el uso de diccionarios de palabras.

Esta información la hemos mostrado tanto mediante histogramas, gráficas y nube de palabras para poder hacer una comparación sobre los resultados obtenidos de forma cómoda.

El tema que hemos seleccionado para analizar ha sido el Brexit. Como hemos ido viendo en las figuras mostradas a lo largo del trabajo, el sentimiento que predomina es joy (divertido): esto nos ha sorprendido ya que nosotros esperábamos que el sentimiento predominante fuese negativo como miedo o tristeza.

En la nube de palabras hemos podido ver las palabras que se relacionan con los sentimientos: en el sentimiento disgust destacan las palabras coming, economy, sexism... mientras que en el sentimiento fear destacan risk, bank, tech....

Además, otro tipo de información que mostramos es la comparativa de si la gente sube imágenes o añade links en los tweets o no. Se puede decir que, tras realizar diversas pruebas, la tendencia general es que los tweets emitivos por iPhone suelen incorporar más imágenes o links que los emitidos por los otros dispositivos.

Hemos añadido un buscador de Trending Topics por localización, una comparativa de las palabras relevantes según la plataforma usada, buscar hastag a partir de usuarios y obtener los usuarios que más tweetean sobre el hastag.

En conclusión, podemos decir que no hay una tendencia clara, ya que como hemos visto en las votaciones, las opiniones positivas y negativas se encuentran prácticamente en equilibrio. Hemos podido comprobar que hay emociones de todo tipo destacando la diversión, lo cuál parece un tanto sorprendente. Gracias a la nube de palabras hemos podido sacar algunos temas candentes como son las relaciones con Polonia y algunos personajes influyentes del momento como es Corbyn.

## 10. Bibliografía.

Algunos de los enlaces de los que hemos hecho uso son los siguientes:

- Para la conexión a la API Twitter.
- Para análisis de sentimientos.
- Para comparativa iPhone, iPad y Android.

Proporcionamos el enlace a nuestro repositorio en github:  
Enlace a repositorio.

Finalmente, volvemos a proporcionar el enlace a nuestra aplicación en Shiny:  
Enlace a la aplicación Shiny.