

Universidad de Alcalá

Escuela Politécnica Superior

Grado en Ingeniería Informática

Trabajo Fin de Grado

Análisis de datos en redes sociales: Caso de estudio aplicado en
Twitter

Autor: David Márquez Mínguez

Tutor: Juan José Cuadrado Gallego

2021

UNIVERSIDAD DE ALCALÁ
ESCUELA POLITÉCNICA SUPERIOR

Grado en Ingeniería Informática

Trabajo Fin de Grado

Análisis de datos en redes sociales: Caso de estudio aplicado en
Twitter

Autor: David Márquez Mínguez

Tutor: Juan José Cuadrado Gallego

Tribunal:

Presidente:

Vocal 1º:

Vocal 2º:

Fecha de depósito: de de

Agradecimientos

A todos los que la presente vieron y entendieron.

Inicio de las Leyes Orgánicas. Juan Carlos I

Aqui va la parte de agradecimientos.....

Resumen

Resumen..... correo de contacto: David Márquez Mínguez <david.marquez@edu.uah.es>.

Palabras clave: Trabajo fin de /grado, L^AT_EX, soporte de español e inglés, hasta cinco....

Abstract

Abstract..... contact email: David Márquez Mínguez <david.marquez@edu.uah.es>.

Keywords: Bachelor final project , L^AT_EX, English/Spanish support, maximum of five....

Resumen extendido

Con un máximo de cuatro o cinco páginas. Se supone que sólo está definido como obligatorio para los TFGs y PFCs de UAH.

Índice general

Resumen	vii
Abstract	ix
Resumen extendido	xi
Índice general	xiii
Índice de figuras	xv
Índice de tablas	xvii
Índice de listados de código fuente	xix
Índice de algoritmos	xxi
Lista de acrónimos	xxi
Lista de símbolos	xxi
1 Introducción al scraping en la web: Marco teórico	1
1.1 ¿Que es el web scraping?	1
2 Presupuesto	3
Bibliografía	5
Apéndice A Funciones implementadas	7
A.1 Función de extracción de Tweets	8

Índice de figuras

1.1	Funcionamiento de Oauth1a	2
-----	-------------------------------------	---

Índice de tablas

Índice de listados de código fuente

A.1 Extracción de Tweets empleando una funcion en R	8
---	---

Índice de algoritmos

Capítulo 1

Introducción al scraping en la web: Marco teórico

Comienzo del trabajo...aaaa

1.1 ¿Que es el web scraping?

Para acceder a la API de Twitter, como se indica en la documentación de la misma, existen dos métodos de acceso OAuth2 y OAuth1a. El acceso con cada uno de ellos dependerá del tipo de información que se desee extraer [1]. Como veremos a continuación, el método de acceso elegido será OAuth1a.

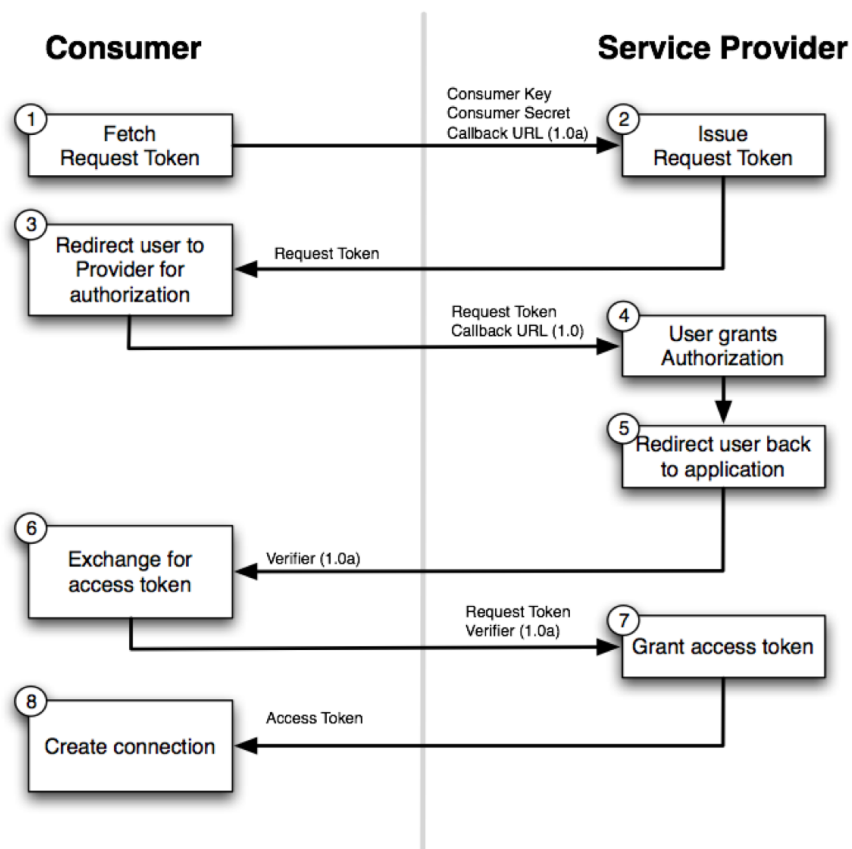


Figura 1.1: Funcionamiento de OAuth1a

Capítulo 2

Presupuesto

Blah, blah, blah.

Bibliografía

- [1] D. Spring, “Service provider ’connect’ framework,” <https://docs.spring.io/spring-social/docs/1.0.0.RC1/reference/html/serviceprovider.html> [Ultimo acceso 16/marzo/2021].

Apéndice A

Funciones implementadas

A.1 Función de extracción de Tweets

Listado A.1: Extracción de Tweets empleando una funcion en R

```
extraccion_tweets <- function(usuario, maxtweets = 100, archivoSalida= NULL){

  #Se crea el nombre de archivo por defecto
  if(is.null(archivoSalida)){
    archivoSalida <- paste0("datos_tweets_", usuario, ".csv")
  }

  #Se comprueba si el archivo csv existe o no
  if(!(archivoSalida %in% list.files())){
    datos_new <- searchTwitter(usuario, n = maxtweets, exclude:retweets)
    datos_new_df <- twListToDF(datos_new)
    write.csv(datos_new_df, archivoSalida)
  }else{
    #Obtengo los datos antiguos
    datos_old <- read.csv(file = archivoSalida)

    #Calculo el id del nuevo tweet a obtener
    ultimo_id <- tail(datos_old, 1)["id"] %>% pull()
    ultimo_id = ultimo_id + 1

    datos_old <- map_if(.x = datos_old, .p = is.numeric, .f = as.character)

    datos_new <- searchTwitter(usuario, n = maxtweets, maxID = ultimo_id, exclude:retweets)
    datos_new <- map_if(.x = datos_new, .p = is.numeric, .f = as.character)

    datos_new_df <- as.data.frame(datos_new)
    datos_old_df <- as.data.frame(datos_old)

    #Se concatenan los nuevos tweets con los antiguos
    datos_concatenados <- bind_rows(datos_old_df, datos_new_df)

    write.csv(datos_concatenados, archivoSalida)
  }
}
```

El objetivo fundamental de esta función es extraer los tweets publicados por un usuario y almacenarlos en un archivo csv. En caso de que exista un archivo con el mismo nombre, se lee y se concatena el nuevo contenido con el antiguo. Los argumentos de entrada, son los siguientes:

1. usuario: representa el identificador del usuario de Twitter.
2. maxtweets: cantidad de tweets que se van a recuperar.
3. archivoSalida: nombre del fichero de salida.

Se debe tener cuidado de no recuperar tweets repetidos, para ello se ha creado una variable que almacena el id del ultimo tweet. Con esto cada vez que se quieran recuperar nuevos tweets, se aumentará en uno dicha variable y se procederá como hasta ahora.

Universidad de Alcalá
Escuela Politécnica Superior



ESCUELA POLITECNICA
SUPERIOR



Universidad
de Alcalá