

# **Manual de usuario del proyecto de generar una nube de palabras en Google Colab de la página Stack Overflow. El documento debe estar hecho en $\text{\LaTeX}$**

Márquez Pincay Jean Carlos

31 de enero de 2021

## **1.Introducción**

El presente documento se redacta con el objetivo de que sirva como guía para el usuario de cómo funciona el software que genera una nube de palabras de la página Stack Overflow en Google Colab. El documento está realizado en  $\text{\LaTeX}$ , el entorno de trabajo del software es Google Colab y su lenguaje de programación es Python.

La nube de palabras permite observar que palabras son tendencia de un texto específico como el de una página web, artículos científicos, tesis, entre otras. Estas palabras son representadas en una sola imagen donde la palabra con mayor frecuencia tiene un mayor tamaño de letra y este tamaño se va reduciendo entre menos frecuencia tenga, como resultado permite al usuario hacer un análisis visual de esta frecuencia para tomar alguna decisión.

Google Colab es un servicio cloud basado en el notebook de Jupyter que permite el uso de los servidores de Google para que los desarrolladores puedan ejecutar código del lenguaje de programación de Python de forma gratuita. Para acceder a este entorno de trabajo se debe ingresar en su página principal, pero para guardar y compartir se necesita una cuenta de Google.

A continuación, se presenta el funcionamiento del software en código y ejecución para que el proyecto tenga éxito en su implementación.

## Información para el uso de la documentación

Este archivo se encuentra bajo una licencia Creative Commons Reconocimiento-CompartirIgual (CC BY-SA). Se permite el uso comercial de la obra y de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.



Versión actual: 31/01/2021

Puedes obtener la última versión de la original disponible de este documento en:

<https://github.com/MarquezJean1/Proyecto-Parcial-I>

## Conceptos de las operaciones

```
from os import path
from PIL import Image
import numpy as np
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS
```

Figura 1: Librerías wordcloud

En la figura 1 se muestra las librerías que permiten cargar los métodos para crear la nube de palabras.

```
import bs4 as bs
from bs4 import BeautifulSoup
import urllib.request
import os
```

Figura 2: Librerías BeautifulSoup

En la figura 2 se muestra las librerías que permiten obtener los métodos para descargar la información de una página web.

```
def leerId():                                #lee numero enteros
    while True:
        try:
            entrada = int(input("Ingrese el ID: "))
            return str(entrada)                #Convierto de int a string
        except ValueError:                     #encaso si sale un exception
            print("El id es solo numero entero")
```

Figura 3: Método validar enteros

En la figura 3 se observa el método que permite validar la información que ingresa el usuario, este método tiene un bucle que capta la información, si esta información genera un error por que no es un número el try-except capta este error y el bucle vuelve a solicitar el ID cuando el ID es correcto se retorna el valor y se rompe el bucle.

```

idUser=leerId()
url = 'https://es.stackoverflow.com/users/' + idUser+'/?tab=tags' #crea el url con el id
try:
    sauce=urllib.request.urlopen(url).read() # Cuando no se encuentre el usuario
    soup = BeautifulSoup(sauce, 'html.parser') #decarga información de la página
    print('Url del usuario: '+url) #transforma la información a html
                                #presenta la url para ir a la página

```

Figura 4: Primera etapa del código

En la figura 4 se observa la primera parte del programa donde se solicita la Id luego se concatena la información para generar el Url, como también el url ingresado puede crear un error al no encontrar al usuario se crea un nuevo try-except, el software procede a descargar la información y a transformarlo a texto html, por último imprime la dirección del usuario.

```

txt=soup.get_text()
file = open("texto.txt", "w") #guarda el texto
file.write(txt) # crea un archivo
file.close() # escribe el archivo
            #cierra el flujo

```

Figura 5: Segunda etapa del código

En la figura 5 se observa como se crea un archivo de texto que almacena la información con el métodos file.write(arg) y como se cierra este flujo con file.close().

```

with open('texto.txt', encoding='utf-8') as f: #recupera el archivo y lo cambio a utf-8
    document = ''.join(f.readlines()) # lee el archivo y guarda en document
#print (document) # para ver el archivo
wordcloud = WordCloud().generate(document) #instancia wordcloud
plt.imshow(wordcloud, interpolation='bilinear') #para dar características
plt.axis("off")
except: #en caso aparezca un error
    print("Lo sentimos no se encontro el usuario.")

```

Figura 6: Etapa final del código

En la figura 6 se observa como se recupera el archivo de texto con el método **with open (arg,encoding='utf-8') as f** donde este lo codifica a texto legible para la función wordcloud y lo guarda a la variable f. Se crea una variable Wordcloud y se agrega la información que tiene la variable document luego se le da forma a la imagen con los métodos y se agrega la instrucción en caso que no encuentre al usuario.

### Procedimientos

El software tiene acceso público necesita tener una cuenta de Google Colab para ejecutar el código e ingresar por el link del proyecto que comparte el desarrollador:

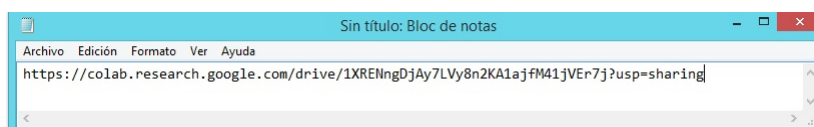


Figura 7: Link del proyecto

[colab.research.google.com/drive/1XRENngDjAy7LVy8n2KA1ajfM41jVEr7j?usp=sharing](https://colab.research.google.com/drive/1XRENngDjAy7LVy8n2KA1ajfM41jVEr7j?usp=sharing)

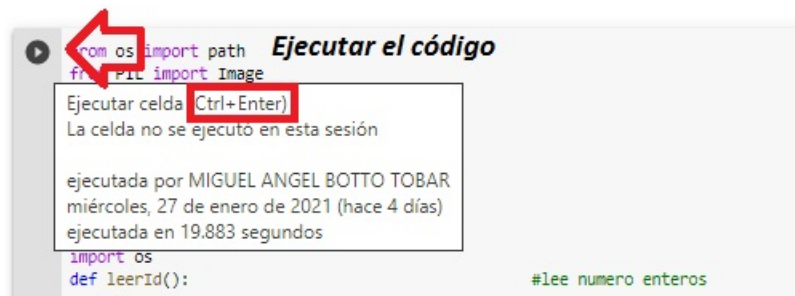


Figura 8: Ejecutar código

A continuación se presenta la página con el código ,para ejecutarlo puede darle click en el botón señalado en la figura 8 o ejecutar el comando **CTRL+ENTER**

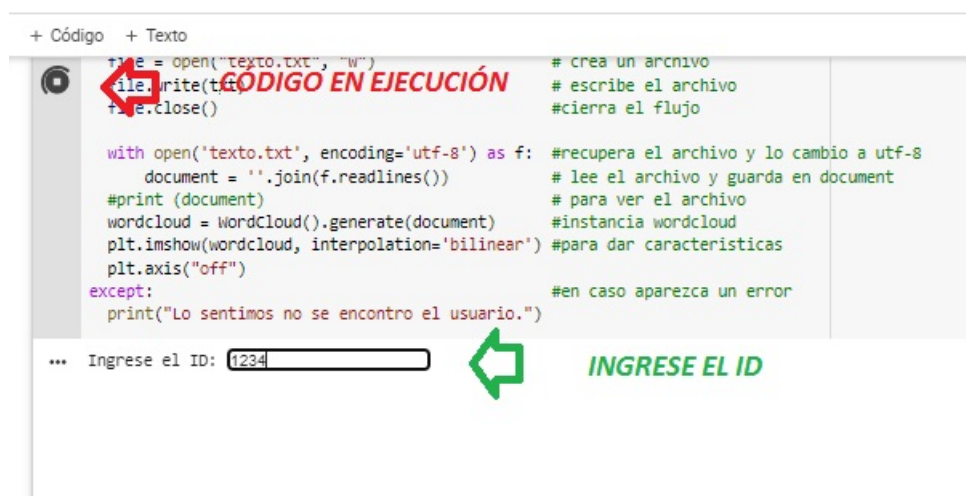


Figura 9: código en ejecución

Como se observa en la imagen 9 cuando el código esta en ejecución el simbolo cambia, si le da click de nuevo en ese botón o usa el comando **CTRL+M+I** el programa se detiene y toma el color de rojo. Debe ingresar el Id como se muestra en la figura 9 y presionar enter. Como resultado mostrara la nube de palabras de esa página.

#### Información sobre los comandos de software

- Se puede ejecutar el código con el comando **CTRL+ENTER**.
- Cuando este ejecutando el programa puede detenerlo con el comando **CTRL+M+I**.

#### Mensajes de error y resolución de problemas

El primer error que puede pasar es cuando se detiene el programa y se lo soluciona ejecutandolo de nuevo:

```
KeyboardInterrupt                                Traceback (most recent call last)
/usr/local/lib/python3.6/dist-packages/ipykernel/kernelbase.py in _input_request(self, prompt, ident, parent, password)
    728         try:
--> 729             ident, reply = self.session.recv(self.stdin_socket, 0)
    730         except Exception:

----- 6 frames -----
zmq/backend/cython/socket.pyx in zmq.backend.cython.socket.Socket.recv()
zmq/backend/cython/socket.pyx in zmq.backend.cython.socket.Socket.recv()
zmq/backend/cython/socket.pyx in zmq.backend.cython.socket._recv_copy()

KeyboardInterrupt:

During handling of the above exception, another exception occurred:

KeyboardInterrupt                                Traceback (most recent call last)
/usr/local/lib/python3.6/dist-packages/ipykernel/kernelbase.py in _input_request(self, prompt, ident, parent, password)
    732         except KeyboardInterrupt:
    733             # re-raise KeyboardInterrupt, to truncate traceback
--> 734             raise KeyboardInterrupt
    735         else:
    736             break

KeyboardInterrupt:
```

Figura 10: Interrupción del proceso

Los siguientes errores es cuando se ingresa letras o caracteres, se lo solucionó con un método para validar el ingreso de la información como se muestra en la figura 3 y cuando no se encuentra el usuario se lo solucionó con un try-except para atrapar el error como se muestra en la figura 4 y 6. Cada error muestra un mensaje como se observa en la figura 11

```
Ingrese el ID: usuario
El id es solo numero entero
Ingrese el ID: 123456789
Lo sentimos no se encontro el usuario.
```

Figura 11: Error ingresar letras y no encontrar el usuario

## Glosario

Sin glosario.

## Referencias

- [1] Luis Ubaldo Godínez.Jun-2011, Estándar IEEE Std 1063-2001, para documentación de usuarios, <http://documentandosistemas.blogspot.com/2011/06/el-estandar-ieee-std-1063-2001.html>.
- [2] Luis Ubaldo Godínez.Jul-2013, ¿Cómo escribir un manual de usuario?, <http://documentandosistemas.blogspot.com/2013/07/guia-basica-para-documentadores-como.html>.