

Orden Digital

---

## Web Data

---

**Elaborado por:** Aarón Josué Sibaja Villalobos

Junio, 2017

# Índice

<b>1. Descripción:</b>	<b>4</b>
<b>2. Consideraciones previas:</b>	<b>4</b>
<b>3. Módulos del Proyecto:</b>	<b>4</b>
3.1. Data . . . . .	5
3.2. Robot . . . . .	5
3.3. Twitter . . . . .	5
<b>A. Cronograma</b>	<b>5</b>

Índice de figuras

1. Cronograma de Actividades. . . . . 5

## 1. Descripción:

El presente documento detalla la funcionalidad del proyecto Web Data, cuyo propósito es extraer metadata de las redes sociales asociados a una búsqueda (palabras o conjunto) dada por el usuario. Esto con el propósito de hacer posteriormente una análisis exploratorio o predictivo con los datos.

El proyecto se encuentra en el lenguaje de programación *Python*. Esto con el propósito de poder sacar ventaja de sus herramientas para el desarrollo de *Web Mining*.

## 2. Consideraciones previas:

Para hacer uso del correcto del programa se debe de hacer una previa instalación del lenguaje de programación Python. Es preferible hacer uso de una version 2.7.XX correspondiente a su última actualización, esto para garantizar su estabilidad. Se recomienda descargarlo de la página oficial:

<https://www.python.org/downloads/>

Seguidamente que se tiene instalado no está de más verificar la instalación y versión de *Python*. Mediante el gestor de paquetes nativo de *Python* `pip` se deben de instalar librerías para el manejo de Twitter:

- `pip install tweepy`
- `pip install twitterstream`

Luego se debe de informar a Twitter que se va a hacer uso de su plataforma, mediante su página oficial de desarrollo Twitter Application Management:

<https://apps.twitter.com/>

Después de crear una cuenta, se debe crear una nueva aplicación. Hay que insertar, un nombre para la aplicación, una breve descripción y algún website, para este caso se utilizo el repositorio de Github. Lo anterior se hace con el motivo de generar *Key y Access Tokens*, credenciales necesarias para extraer los datos desde Twitter.

## 3. Módulos del Proyecto:

El proyecto se divide en tres partes principales:

### **3.1. Data**

Contiene los archivos referentes el manejo de la manipulación de los datos. Se encuentra la opción de guardado a una base de datos MySQL o como salida en un archivo csv.

### **3.2. Robot**

Contiene los módulos para crear demonio que se ejecute a conveniencia. Esto con el propósito de establecer intervalos de extracción personalizados dependiendo de las exigencias que se presenten.

### **3.3. Twitter**

Contiene lo códigos para la extracción de los datos en la red social. Contiene las respectivas credenciales configuradas específicamente para la aplicación como se detallo anteriormente.



## A. Cronograma

Proyecto Web Data

Nombre de tarea	Duración	Comienzo	Fin
Inicio web data	6 horas	mié 15/3/17	mié 15/3/17
Estudio Twitter	2 horas	jue 16/3/17	jue 16/3/17
Estudio Facebook	4 horas	vie 17/3/17	vie 17/3/17
Twitter Stream	6 horas	sáb 18/3/17	sáb 18/3/17
Resultados Búsqueda	4 horas	dom 19/3/17	dom 19/3/17
Módulo SQL	4 horas	lun 20/3/17	lun 20/3/17
CSV	3 horas	mié 22/3/17	mié 22/3/17
Manejo CSV	4 horas	vie 24/3/17	vie 24/3/17
Demonio	5 horas	dom 26/3/17	dom 26/3/17
Demonio	4 horas	mié 29/3/17	mié 29/3/17
Resultados Búsqueda	4 horas	vie 31/3/17	vie 31/3/17
Consultas	3 horas	sáb 1/4/17	sáb 1/4/17
Corrección Bug Búsqueda	3 horas	dom 2/4/17	dom 2/4/17
Corrección Bug Búsqueda	4 horas	mié 5/4/17	mié 5/4/17
Manejo Session Twitter	5 horas	vie 7/4/17	vie 7/4/17
Inicio Modelo BD	2 horas	jue 6/4/17	jue 6/4/17