

**UNIVERSIDAD DE GUAYAQUIL**

**FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS**

**INGENIERIA EN SOFTWARE**

**MATERIA:**

PROCESOS DE SOFTWARE

**DOCENTE:**

Msg. Miguel Angel Botto Tobar

**PROYECTO PRIMER PARCIAL**

Generador de nube de palabras a partir de archivos de bases de datos de los usuarios de StackOverflow en español.

**INTEGRANTES:**

* CANTOS PINTO JAMES DOUGLAS
* MINA AREVALO JEFFERSON PERLAZA

**CURSO: SOF-S-MA-3-2**

**Introducción:**

Una nube de palabras o nube de etiquetas es una representación visual de las palabras que conforman un texto, en donde el tamaño es mayor para las palabras que aparecen con más frecuencia.

Una nube de palabras o nube de etiquetas es una representación visual de las palabras que conforman un texto, en donde el tamaño es mayor para las palabras que aparecen con mayor frecuencia, el proyecto consiste en generar una nube de palabras a partir de la base de datos de etiquetas de la plataforma <https://es.stackoverflow.com/> los cuales contienen los nombres de las etiquetas que necesitamos para generar la nube de palabras. Hay que generar la nube de palabras con los nombres de las etiquetas y basándose en todas las palabras que mencionan los usuarios, y la más mencionada debe ser la más grande.

Para este proyecto utilizaremos el lenguaje Pyhton,la versión 3.7.

**Primero que todo tuvimos que buscar nuestros archivos xml y de usuario de StarckOverflow en la base de datos.**

**Herramienta para búsqueda de archivos xml de base de datos:**

<https://archive.org/details/documentation-dump.7z>

**Dirección del archivo de la base de datos de usuarios de StackOverflow en español:**

<https://archive.org/details/stackexchange>

Debemos descargar la base de datos tags.xml que se encuentra dentro de la carpeta comprimida es.StackOverflow.con en formato xml.

Imagen que contiene captura de pantalla

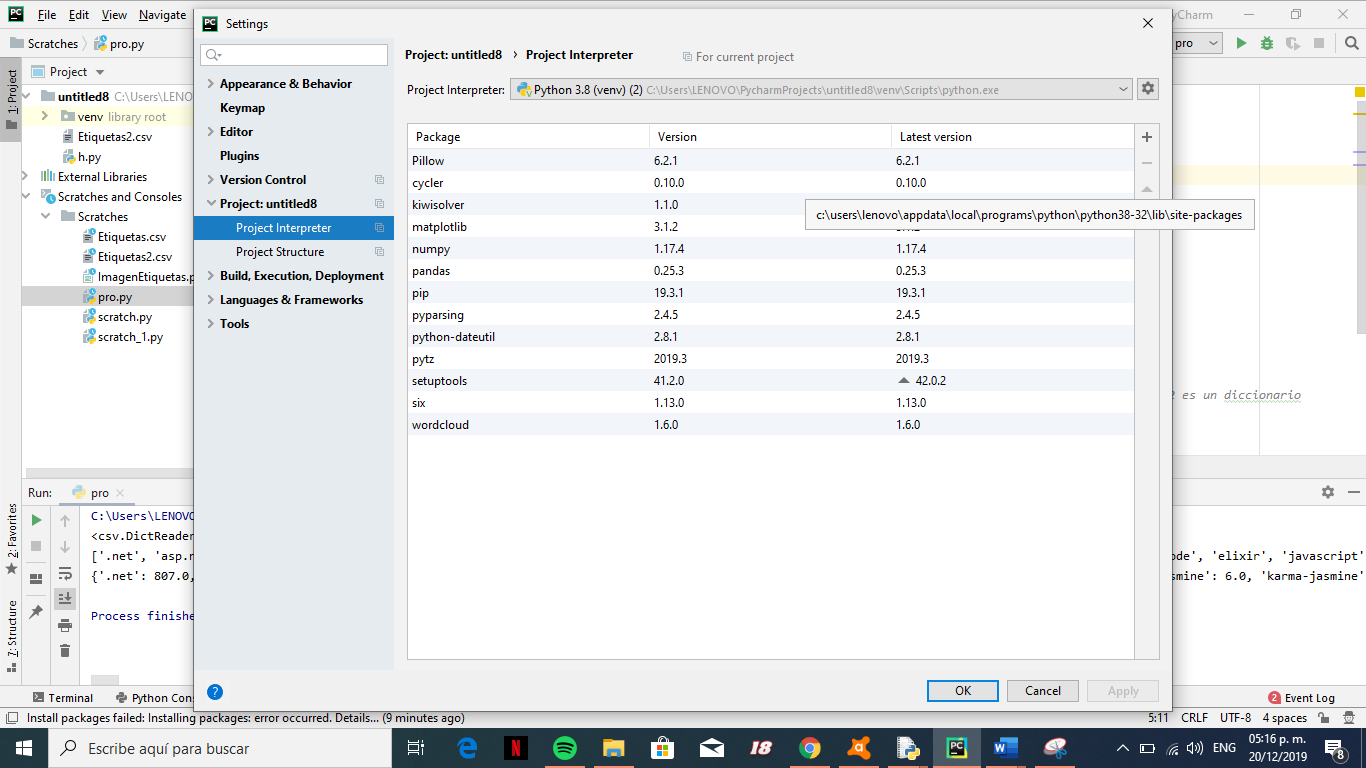
Descripción generada automáticamente

Luego convertimos nuestro archivo xml a formato cvs.

Imagen que contiene captura de pantalla

Descripción generada automáticamente

Una vez ya tenemos la información de la base de datos, nos adentramos en la programación, debido a que la solución que pretendemos implementar requiere un algoritmo de renderizado gráfico a demás de manipular texto y generar datos estadísticos, para esto solo usaremos WordClouds como librería externa



Una vez descargadas las librerías externas ya podemos comenzar a codificar empezando por importar las librerías Wordcloud que es la que nos va a permitir crear la nube de palabras, pudiendo medicar su tamaño y color ( apariencia)

Para crear nuestro diccionario como primer paso le damos a csv2 como parámetro nuestro archivo Etiquetas2.csv.

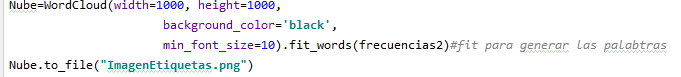
Luego importamos nuestro csv,asignameros el método DictReader que sirve para leer sabiendo que valor es cada columna,luego la imprimimos print(reader).

Con el bucle for decimos que para cada línea que tenga reader,nuestro método “append”agrega cada nombre.

Imagen que contiene captura de pantalla

Descripción generada automáticamente

Como penultimo paso generamos la nube de palabras mediante el método Word.cloud usando como parámetro el diccionario de palabras y frecuencias, nuestro método es Nube.to\_file y nuestro parámetro es (“ImagenEtiquetas.png”) para generar un archivo con el nombre especificado como parámetro.



**Como último paso ejecutamos nuestro programa**Imagen que contiene texto, periódico

Descripción generada automáticamente.