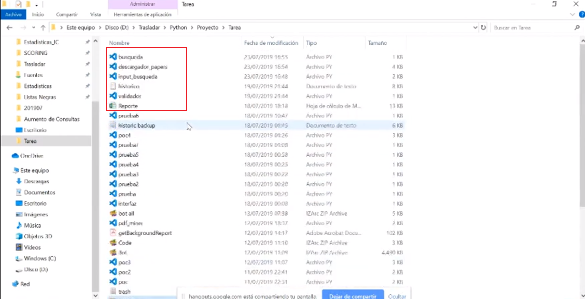
MANUAL DE FUNCIONAMIENTO Y EJECUCIÓN DE BOT EN PYTHON

**Funcionamiento**

El programa ha sido creado en Python, para esto se utilizó principalmente las librerías SELENIUM para la automatización y PYPDF2 para la captura de metadatos de los archivos pdf descargados.

El programa se creó y codificó en distintos módulos que veremos a continuación.



Acá se muestra los archivos .py utilizados para el funcionamiento del programa:

* **Input\_busqueda**; este archivo es la interfaz que captura los parámetros de búsqueda. Es modificable según los parámetros que sean conveniente en la búsqueda de los repositorios.
* **Búsqueda y validador**; estos archivos son necesarios para almacenar los parámetros de búsqueda e iniciar la búsqueda.
* **Descargador\_papers**; tal y como lo dice su nombre contiene el código para la descarga de los archivos.

Otros:

* **Historico**; archivo que registra historial y datos de los archivos descargados.
* **Reporte**; archivo Excel que se actualiza a través del histórico, en el cual se ordena y almacena los registros deseados de los archivos descargados.

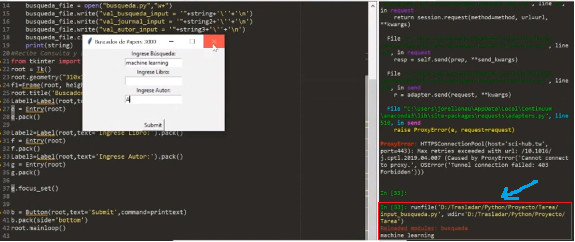
Nota: Es necesario tener descargado el siguiente ejecutable.



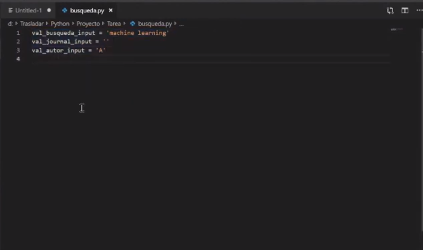
Link: <http://chromedriver.chromium.org/downloads?fbclid=IwAR3T1DQNOgVnpJlMGEr6u_CGZr-0CgMm8TVdegZA-rlZUBruCOnWg9gaEXY>

**Ejecución**

Se explicará brevemente como se ejecuta la búsqueda de acorde a los archivos utilizados y los resultados de la búsqueda.

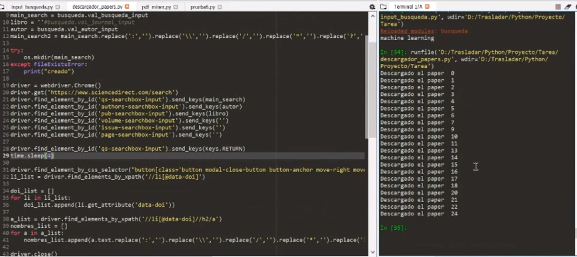


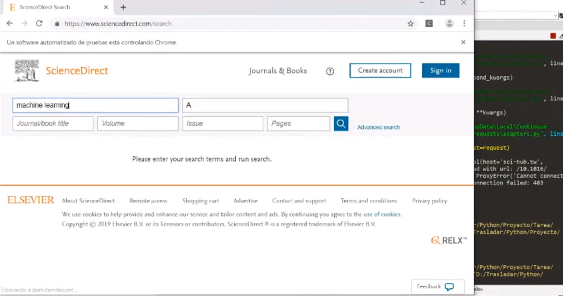
En la imagen superior vemos una pequeña interfaz para la captura de los parámetros de búsqueda; en el ejemplo vemos que se llena los datos de búsqueda de tema y autor (inicial o parte de una cadena de referencia para filtrar por nombre), se ingresa “machine learning” y la inicial ‘A’ para luego presionar Submit. Inmediatamente en la parte inferior derecha vemos que el primer paso es almacenar los parámetros de búsqueda.



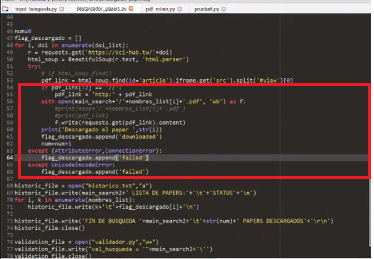
En la imagen superior vemos que el archivo “busqueda.py” ha capturado los datos ingresados.

A continuación, en la imagen inferior vemos el archivo “descargador\_paper.py” el cual primero valida los datos de búsqueda según los datos capturados por el archivo “busqueda.py”. Luego inicia la búsqueda en el repositorio web automáticamente, para finalmente iniciar el proceso de descarga el cual vemos en la parte derecha de la imagen.



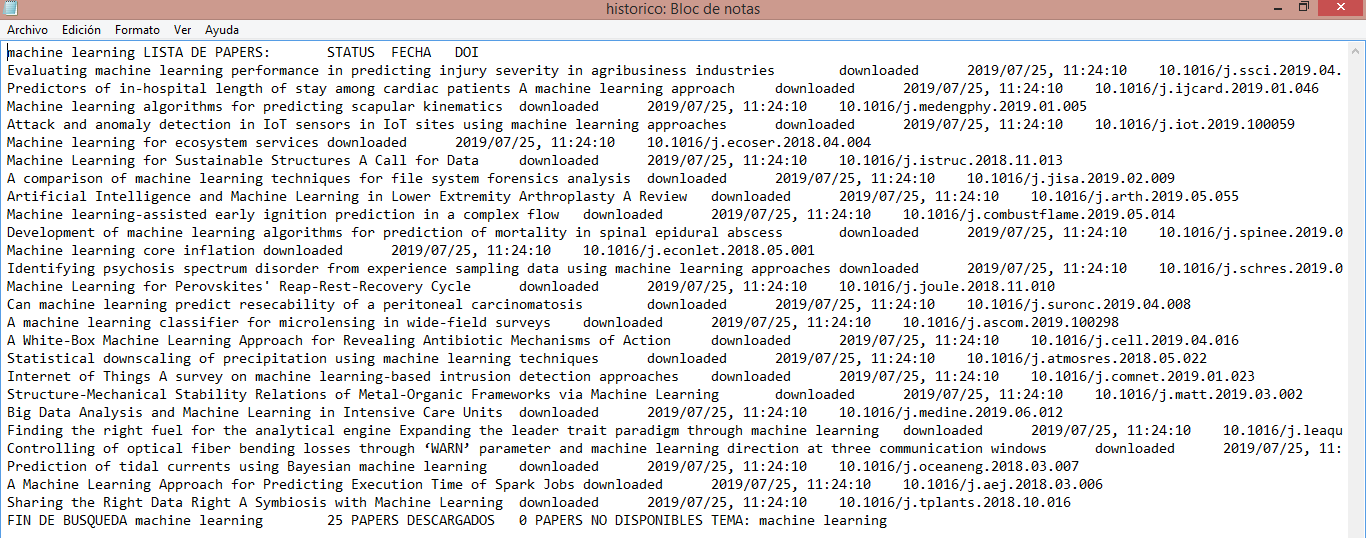


En caso error de descarga en algún archivo el “descargador\_paper.py” lo captura para luego saltarse ese error y continuar con la descarga. Ver imagen inferior.

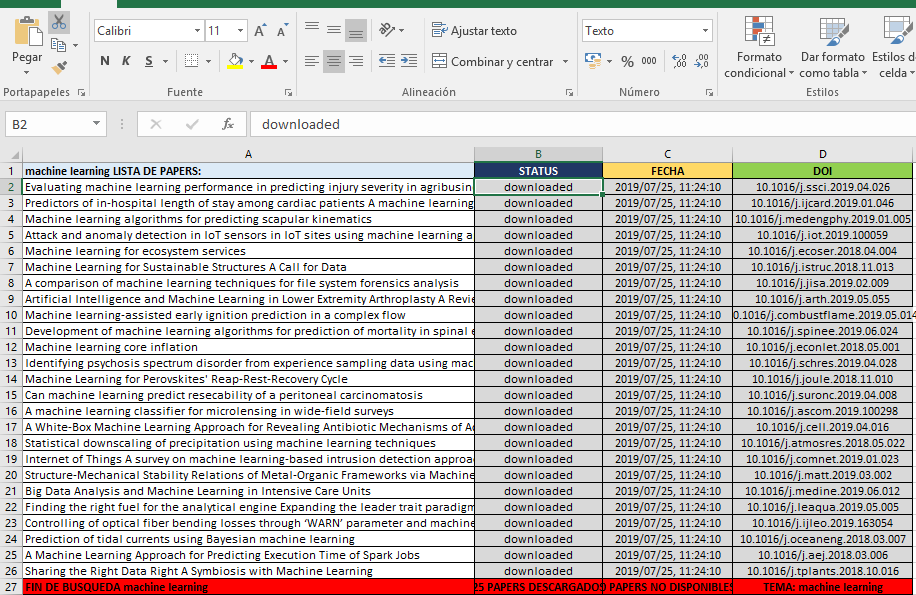


**Resultados**

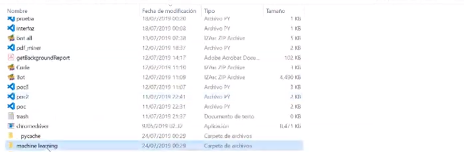
Primero vemos el archivo histórico que registra los datos de los papers descargados para luego generar un Excel con los registros almacenados.

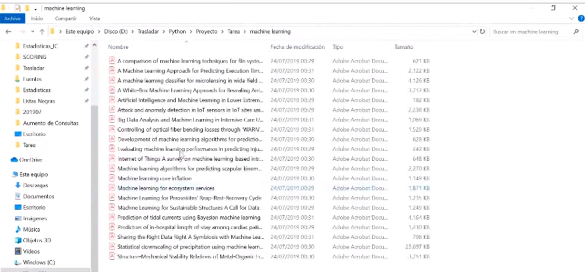


A continuación, veremos archivo “Reporte”, que es el Excel que se ha generado con los datos más relevantes de los papers descargados.



Finalmente visualizamos que se crea una carpeta con el nombre correspondiente “machine learning” al tema de descarga que contiene los archivos descargados.





**UNIVERSIDAD NACIONAL DE INGENIERÍA**

**Curso: Administración del Conocimiento**

**Docente: Vilcapoma Escurra, Edgar**

**Creadores:**

* **Miller Conyas, Jorge Ennrique**
* **Orellana Ugaz, Julio Cesar**
* **Limaquispe Huaman, Miguel Angel**