

Práctica 1. Web Scraping

El objetivo de esta actividad será la creación de un dataset a partir de los datos contenidos en una web.

1. Contexto.

El conocimiento de la información de los accionistas de las compañías del Ecuador, es útil para realizar campañas de prospección y venta de productos, así como también identificar el capital social de cada compañía.

Se seleccionó la web de la Superintendencia de Compañías, Valores y Seguros del Ecuador por cuanto es la institución pública, que vigila y controla la organización, actividades, funcionamiento, disolución y liquidación de las compañías, por lo que es la única fuente de datos donde se puede obtener la información de los accionistas.

2. Título Dataset.

Accionista_Compañía.csv

3. Descripción del Dataset.

El Dataset generado contiene 9 columnas y 22 registros de accionistas correspondientes a 4 empresas seleccionadas.

4. Representación Gráfica.



5. Contenido.

Los campos que se incluyen en el dataset son los siguientes:

numero. Identifica el número de accionista registrado de la empresa, es un número entero secuencial que inicia en 1.

identificación. Corresponde al número identificación del accionista de la compañía, es un campo de texto.

accionista. Apellidos y Nombres si el accionista es persona natural o razón social si el accionista es una compañía, es un campo de texto.

nacionalidad. Nacionalidad del accionista, es un campo de texto.

tipo_inversion. Tipo de inversión del accionista, puede tomar uno de los siguientes valores: EXT. DIRECTA, SUBREGIONAL, NACIONAL.

capital. Monto en dólares de la inversión (acciones) realizada por el accionista, es un número flotante.

restricción. Indicador de si el accionista tiene o no alguna restricción, puede tomar los valores S o N.

expediente. Número de expediente (identificador de la compañía) definido en la Superintendencia de Compañías.

empresa. Nombre de la compañía.

El campo **expediente** se obtuvo del directorio de compañías que está disponible en la web de la Superintendencia de Compañías del Ecuador, el resto de campos se obtuvieron a partir del proceso de web scraping implementado.

6. Agradecimientos.

Agradecimientos a la Superintendencia de Compañías, Valores y Seguros del Ecuador que es una institución pública que controla, vigila y promueve el mercado de valores, el sector societario y de seguros, mediante sistemas de regulación y servicios, contribuyendo al desarrollo confiable y transparente de la actividad empresarial en el país.

7. Inspiración.

El conjunto de datos Accionista_Compañía, contiene los accionistas de una muestra de compañías del Ecuador con los valores en dólares de sus inversiones (acciones) en cada compañía, esta información es útil para diferentes propósitos, prospección y ventas de productos, cumplir con normas como “Conozca a su cliente”, etc.

En esta práctica se pretende responder a la pregunta. ¿Cuál es el capital social de cada una de las compañías seleccionadas.

8. Licencia.

La licencia para el dataset generado es: Database released under Open Database License, por cuanto la licencia abierta de bases de datos permite a los usuarios compartir sus datos con libertad y sin temor a los derechos de autor o cuestiones de propiedad, ya que la información del dataset se obtiene de fuentes de información pública.

9. Código.

El código con el que se ha generado el dataset *Accionista_Compañía*, se ha generado en Python y se encuentra alojado en el repositorio Github que contiene este documento.

10. Dataset.

El dataset en formato CSV está publicado en Zenodo con el siguiente DOI:

10.5281/zenodo.4679915

<https://zenodo.org/search?page=1&size=20&q=4679915>

Contribuciones		Firma
Investigación previa	LAP, JLM	
Redacción de las respuestas	LAP, JLM	
Desarrollo código	LAP, JLM	