

WebScraping con Python

Practica 1

Descripción

Esta práctica tiene como objetivo aplicar técnicas de WebScraping con python 2.7 para poder extraer datos de la Wikipedia de las provincias del ecuador con datos de cada una y generar como resultado un archivo .csv con los datos obtenidos.

Miembros del equipo

Jennifer Samaniego

Repositorio: <https://github.com/jenn1991/TipologiaPrac1UOC>

Características del Dataset

Los datos que se obtienen son las provincias del ecuador con el número de habitantes, extensión y densidad.

1. **Título del dataset.** Provincias de Ecuador.
2. **Subtítulo del dataset.** Características por provincia.
3. **Imagen.**



Figura 1 Mapa del Ecuador

Autoría: Wikipedia TUBS [CC BY-SA 2.5 (<https://creativecommons.org/licenses/by-sa/2.5>)], via Wikimedia Commons

4. Contexto.

¿Cuál es la materia del conjunto de datos?

Los datos que se están obteniendo están dentro de la categoría del análisis geográfico de un país en específico. Cuando es necesario conocer la distribución geográfica de un país lo podríamos realizar de forma manual, pero tocaría invertir mucho tiempo en revisar las webs cada cierto tiempo para poder conocer si ha cambiado sus diferentes características; es por ello que estoy

utilizando webscraping generando un script en Python el cual extraerá el número de habitantes, extensión y densidad de cada provincia.

5. Contenido.

¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

Los datos tienen su última actualización el día 21 de octubre del 2018 y fueron obtenidos el 11 de noviembre del 2018 se recogieron con la ejecución del script generado en Python los campos que se recogieron son:

- Provincia
- Habitantes
- Extensión
- Densidad

El script generado sirve para poder consultar el resto de campos de las otras tablas si se consideraran necesarias.

6. Agradecimientos.

¿Quién es propietario del conjunto de datos? Includ citas de investigación o análisis anteriores.

Agradezco a Wikipedia por su información la misma se basa en otros documentos; esta web solicita que se realice el reconocimiento de su información adjunto el reconocimiento solicitado: https://es.wikipedia.org/wiki/Provincias_de_Ecuador.

7. Inspiración.

¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?

Me parece interesante primero por ser Ecuatoriana y con este pequeño script pueda consultar ciertos datos de mi país y conocerlo de forma general, en un futuro se podría hacer un script más grande que permita unir la parte turística de cada provincia con los datos generales.

8. Licencia.

Seleccionad una de estas licencias y decid porqué la habéis seleccionado:

La licencia que se escogió es **CC BY-NC-SA 4.0**, para que el código pueda ser modificado y redistribuido siempre y cuando se mencione el creador original, no se utilice para propósitos comerciales y se redistribuyan las modificaciones sobre la misma licencia. De esta manera otros usuarios pueden continuar expandiendo el script para la búsqueda de información geográfica de otros países.

9. Código:

Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset

El código se encuentra en la carpeta SRC en el repositorio.

10. Dataset:

Dataset en formato CSV se encuentra en la carpeta CSV en el repositorio.