**Propuesta de Curso para Retribución Social**

Rocío Cruz Lathrop

Herramientas para la recolección de información de texto para la construcción de Bases de Datos

El objetivo del curso es compartir herramientas para la creación de variables cuantitativas a partir de información cualitativa como archivos pdf y/o disponibles en internet. Para esto comparto dos diferentes estrategias de búsqueda sistemática de información cualitativa para el desarrollo de bases de datos. La primera es la técnica de web scraping con clave API y la segunda es búsqueda masiva de archivos con Power Shell de Windows.

Sesión 1

El objetivo de esta sesión es revisar definiciones de variables, su operacionalización y fuentes potenciales de información.

Sesión 2

El objetivo de este módulo es dar a conocer la técnica de web scraping en Google News con Python para la búsqueda de información.

Sesión 3

En el cuarto módulo utilizaremos la búsqueda de información en textos con la consola de Power Shell en Windows.

Como metodología en cada módulo se verán ejercicios con casos prácticos con bases de datos abiertas al público y originales utilizadas en mis proyectos.

Es importante señalar que esto está basado en el aprendizaje autodidacta de estrategias utilizando tutoriales de internet y complementando su comprensión con el uso de Chat GPT.

Contenido

1. Revisión de variables para su operacionalización

El objetivo de esta sesión es la operacionalización de una variable y su diagnóstico inicial de posibles fuentes de información a las que puedo tener acceso.

En esta sesión se analizarán ejemplos de variables operacionalizadas y se desarrollará un ejemplo utilizando la tabla de que propongo a continuación.

Ejemplo

Appendix 3.1, Coding Rules for Political Regimes (pp. 297-299), Democracies and dictatorships in Latin America Emergence, Survival, and Fall (Mainwaring & Pérez-Liñán, 2013)

Tabla para ejercicio

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre de variable | Definición | Ejemplos | Fuentes de información |
|  |  |  |  |
|  |  |  |  |

1. Web Scraping en Google News usando Python

* ¿Conociendo Google Colab?(*Google Colab Tutorial*, s. f.; TensorFlow, 2019)
* Principales librerías. Pandas y Numpy (*10 minutes to pandas — pandas 2.3.1 documentation*, s. f.)
* ¿Qué es y cómo obtengo mi clave Api?
* ¿Cómo desarrollo mi código para webscraping?

Referencias

Khandelwal (2024), Scrape Google News Using Python (Export Data in CSV), disponible en: <https://www.scrapingdog.com/blog/scrape-google-news/>

[Wgunkjcvl](https://github.com/wgunkjcvl), (año), Web Scraping Google News Using Python: A Step-by-Step Guide, disponible en: <https://github.com/wgunkjcvl/SnapShooter/blob/main/practices/Web%20Scraping%20Google%20News%20Using%20Python%3A%20A%20Step-by-Step%20Guide.md>

*10 minutes to pandas—Pandas 2.3.1 documentation*. (s. f.). Recuperado 29 de julio de 2025, de https://pandas.pydata.org/docs/user\_guide/10min.html

*Google Colab Tutorial*. (s. f.). Codefinity. Recuperado 29 de julio de 2025, de https://codefinity.com/ua/blog/Google-Colab-Tutorial?utm\_source=google&utm\_medium=cpc&utm\_campaign=21413601262&utm\_content=&utm\_term=&dki=&gad\_source=2&gad\_campaignid=21409590893&gclid=CjwKCAjw1ozEBhAdEiwAn9qbzZ7q\_\_3bNYJQOAMfhD-6kAj4fV4ikJrKwtuJY6H8-qAUOkoh0dvY1xoC2ikQAvD\_BwE

TensorFlow (Director). (2019, enero 30). *Comienza a usar Google Colaboratory (Coding TensorFlow)* [Video recording]. https://www.youtube.com/watch?v=inN8seMm7UI

1. Búsqueda de palabras o frases con Power Shell de Windows.

Objetivo: Enseñar cómo realizar búsquedas de texto en archivos locales a través de la consola PowerShell en Windows.

- Uso de comandos básicos de búsqueda (Select-String, Get-ChildItem, etc.)  
- Ejercicio guiado con archivos de texto previamente descargados.

Nota

El curso incluye una carpeta compartida con recursos

Se ofrece una cuarta sesión para la revisión de proyectos personales