

Tipología y Ciclo de Vida de los Datos

Práctica 1: Web Scraping

1. Contexto.

En esta práctica se pretende recoger información sobre diferentes indicadores de calidad del aire con el objetivo de determinar el grado de contaminación atmosférica presente en ciertos puntos de la geografía de la comunidad autónoma gallega.

La contaminación atmosférica puede tener su origen en dos factores, en la actividad humana o en la naturaleza (fenómenos naturales que se producen como parte de los ciclos elementales de la materia, vulcanismos, incendios forestales, etc.) Es, en estos últimos, donde centraremos la recogida de información.

La web de la Consellería de Medio Ambiente, Territorio e Vivenda de Galicia, a través de la web MeteoGalicia (www.meteogalicia.gal), proporciona información sobre una serie de indicadores sobre la calidad del aire

SO₂	Dióxido de azufre (Fluorescencia ultravioleta)
NO	Monóxido de nitrógeno (Quimioluminiscencia)
NO₂	Dióxido de nitrógeno (Quimioluminiscencia)
NO_x	Óxidos de nitrógeno totales (como NO ₂) (Quimioluminiscencia)
CO	Monóxido de carbono (Absorción infrarroja)
O₃	Ozono (Absorción ultravioleta)
PM₁₀	Partículas en suspensión (<10µm) (Absorción beta)
PM_{2,5}	Partículas en suspensión (<2.5µm) (Absorción beta)

2. Título para el dataset.

Galician Quality Air Data

3. Descripción del dataset.

El conjunto de datos está formado por datos procedentes del conjunto de estaciones que dispone la Rede Xunta de Galicia, donde se realizan mediciones de los indicadores mencionados en el apartado 1 del documento. Para cada indicador se dispondrá de su valor y de su estado en el momento en el que se realizó la extracción. Dicho estado viene determinado por los siguientes valores:

T	Temporal
M	Equipo en mantenimiento
V	Válido
D	Inválido
C	Calibración de SPAN
Z	Calibración de Cero
N	Nulo

4. Representación gráfica.



5. Contenido.

Para cada estación y fecha y hora en la que se lleve a cabo la medición, el conjunto de datos dispondrá de los valores y estados de cada uno de los indicadores en ese momento. Los campos disponibles:

Campo	Descripción
estacion	Estación a la que pertenecen las medidas de los indicadores
fecha	Fecha y hora en la que se realizó la medición
SO2-valor	Valor del indicador SO ₂
SO2-estado	Estado del valor del indicador SO ₂
NO-valor	Valor del indicador NO

NO-estado	Estado del valor del indicador NO
NO2-valor	Valor del indicador NO ₂
NO2-estado	Estado del valor del indicador NO ₂
NOX-valor	Valor del indicador NO _x
NOX-estado	Estado del valor del indicador NO _x
CO-valor	Valor del indicador CO
CO-estado	Estado del valor del indicador CO
O3-valor	Valor del indicador O ₃
O3-estado	Estado del valor del indicador O ₃
PM10-valor	Valor del indicador PM ₁₀
PM10-estado	Estado del valor del indicador PM ₁₀
PM25-valor	Valor del indicador PM _{2,5}
PM25-estado	Estado del valor del indicador PM _{2,5}

6. Agradecimientos.

La información contenida en la web www.meteogalicia.gal está elaborada a partir de la actividad científica y técnica desarrollada por la Consellería de Medio Ambiente, Territorio e Vivenda perteneciente a la Xunta de Galicia. Toda la información contenida en la web es propiedad intelectual y goza de la protección reconocida por el texto de la Ley de Propiedad Intelectual aprobado por Real Decreto Legislativo 1/1996 del 12 de abril.

La información divulgada en este portal se ofrece gratuitamente a los ciudadanos para que pueda ser utilizada libremente por ellos, con el único compromiso de mencionar expresamente a MeteoGalicia y a la Consellería de Medio Ambiente, Territorio e Vivenda de la Xunta de Galicia, como fuente de esta cada vez que se utilicen para los distintos usos del particular y privado

7. Inspiración.

Este conjunto de datos de datos podría utilizarse para realizar estudios sobre posibles zonas para una mejor ubicación de determinada industria, zonas residenciales, impactos por la construcción de vías de transporte, así como para el estudio de la evolución de la contaminación en dichas zonas o la influencia de las condiciones atmosféricas en dichos indicadores.

8. Licencia.

La licencia escogida para la publicación de este conjunto de datos es CC BY-SA 4.0 License, donde se permite el uso comercial de la obra y de las posibles obras derivadas,

siempre y cuando la distribución se lleva a cabo con una licencia igual a la que regula la obra original.

9. Código.

Ver directorio **source** del repositorio GitHub.

10. Dataset.

Ver directorio **dataset** del repositorio GitHub.

11. Observaciones.

El parámetro “periodo” tiene un comportamiento correcto y si se establece alguno de los valores posibles, el scraper funciona correctamente.

No obstante, no se ha podido completar la parte de modificación del formulario para obtener el rango de fechas ni la estación deseada. Esos parámetros deben dejarse vacíos.

Para realizar una prueba donde se vea el comportamiento erróneo del parámetro “estacion” basta con establecer el valor “Lalin” en dicho parámetro.