

Práctica-Formatos de datos para la Web(I)

Ejercicio 1[5 puntos]

Considerar el archivo emisiones_fuentes_contaminantes_csv_2008 que contiene un conjunto de archivos en formato csv sobre contaminación, se pide realizar funciones en Python que permitan realizar las siguientes acciones:

1. Usando el archivo agua_eprtr_2008_030412 crea un nuevo archivo csv denominado AguaAgrupada.csv el cual debe contener agrupadas las empresas con el tipo de contaminación que producen. Tendrá la estructura:
Contaminación, empresa
CADMIO Y SUS COMPUESTOS, BEFESA ZINC AMOREBIETA
CADMIO Y SUS COMPUESTOS, UROLA KOSTAKO UDAL ELKARTEA
.....
2. Usando el archivo residuos_peligrosos_eprtr_2008_040412 crea un nuevo archivo cvs denominado FrecuenciaResiduos.csv el cual debe contener la frecuencia con la que aparecen las diferentes empresas. Tendrá la estructura:
Jugador,Frecuencia
3. Usando el archivo aire_eprtr_2008_030412 crea un nuevo archivo cvs denominado Contaminantes.csv el cual tenga la información de las 10 empresas que más contaminan. Para ello se debe sumar la cantidad Kg/año de los residuos emitidos por la empresa.

Ejercicio 2[5 puntos]

Considerar el archivo semaforos.txt que contiene información de los semáforos de Málaga en formato json. Se pide realizar funciones en Python que permitan:

- a) Encontrar la frecuencia de información acerca de un semáforo, la cual puede calcularse como el cociente entre el número de ocurrencias de un semáforo y el número de ocurrencias de todos los semáforos. Utiliza el campo ubicación para caracterizar la aparición de un semáforo. Como resultado se debe generar un archivo que tenga en cada línea el nombre de un semáforo seguido de la frecuencia del mismo. Además no puede aparecer líneas con semáforos repetidos. Hay que tener en cuenta lo siguiente:
 - Los datos recuperados que se encuentran en el archivo proporcionado deben convertirse en una lista en python. Para ello se leerá línea a línea el archivo recuperado, de manera que los elementos de la lista serán la información de cada semáforo que se han recuperado.
 - Localiza en cada elemento dónde se encuentra la información de la ubicación.
- b) Encontrar los 10 semáforos que aparecen con mayor frecuencia. Utiliza el campo ubicación para caracterizar la aparición de un semáforo. Como resultado se debe generar un archivo que tenga en cada línea la ubicación del semáforo seguido del número de veces que aparece.

Normas de entrega

- Fecha tope de entrega: 26/10/2016
- La entrega se realizará subiendo un archivo zip con las soluciones realizadas en Python. El archivo tendrá como nombre FormatosWeb(I)_GrupoX donde X será el número de grupo correspondiente.