

Trabajo práctico integrador

Análisis de datos abiertos - Localización de basurales

Seminario de lenguajes opción C

2019

A raíz de un estudio sobre el impacto ambiental de los basurales de la cuenca Matanza - Riachuelo realizado en el año 2017, surgió la necesidad de obtener información estadística sobre los datos que dicho estudio arrojó. Para facilitar su difusión y entendimiento por personas ajenas a la materia específica de medio ambiente, se ha requerido procesar los datos para poder generar un mapa que muestre gráficamente los distintos basurales encontrados.

A tal fin, se tiene a disposición el dataset de basurales de la cuenca Matanza - Riachuelo de 2017 en formato CSV, el cual deberá ser utilizado por un programa que permita cargar ese archivo y realizar una serie de operaciones sobre sus datos, para obtener finalmente la información deseada.

Realizar un programa que, tomando como entrada el dataset en formato CSV y una serie de argumentos, procese la información y permita realizar diferentes operaciones sobre la misma.

El programa deberá admitir los siguientes argumentos:

- **-h|-a|--help|--ayuda:**
Muestra qué argumentos admite el programa con una ayuda breve para cada uno.
- **-l|--leer archivo:**
Carga el archivo CSV de nombre “archivo”.
- **-s|--salida archivo:**
Indica el nombre del archivo donde se generará la secuencia de puntos. Si se indica como archivo “-” o se omite este argumento, la salida se envía a la salida estándar.
- **-m|--municipio:**
Indica que se generará un único punto por cada municipio.

Modo de procesar los parámetros:

- Si se envía el parámetro *-h* el programa muestra una ayuda con todos los parámetros y termina de inmediato (el resto de los parámetros aunque aparezcan antes se ignoran).

- El parámetro *-l* es obligatorio y no se puede repetir, le sigue un nombre de archivo.
- El parámetro *-s* es obligatorio y no se puede repetir, le sigue un nombre de archivo. Si se encuentra *-s* sin *-l* mostrar un mensaje de error y finalizar el programa.
- El parámetro *-m* es opcional, si se envía se ejecutará la funcionalidad de la etapa 2 del trabajo. Si se encuentra *-m* sin *-s* mostrar un mensaje de error y finalizar el programa.
- Especificar *-l* o *-s* sin pasar un nombre de archivo debe mostrar de ser posible un mensaje de error y el programa debe terminar de inmediato.

Si no se puede abrir el archivo de entrada o el de salida, el programa deberá mostrar un error indicando el nombre del archivo que no pudo abrir y terminar el programa de inmediato.

Durante la fase de procesamiento de parámetros y apertura de archivos si se detecta algún error el programa deberá terminar con uno de los siguientes códigos de error (número indicado a la izquierda):

- 65. Si falta un argumento requerido.
- 66. Si falta el nombre de archivo en *-l* o *-s*.
- 67. Si hay argumentos extra o desconocidos.
- 68. Si no se pudo abrir un archivo para lectura.
- 69. Si no se pudo abrir un archivo para escritura.

Agregar los códigos de error que sean necesarios para reportar otros errores posibles, como por ejemplo: archivo CSV no válido o campo muy largo.

Parte 1

Entrega: Semana 24/06/2019

Implementar la funcionalidad necesaria para los argumentos *-l* y *-s* (o sus versiones largas).

A partir del archivo CSV recibido como parámetro de *-l* el programa deberá generar un archivo con puntos para mostrar en mapcustomizer.com usando la opción “Bulk Entry” que muestre los basurales con el siguiente formato:

```
latitud,longitud {tipologia: denominacion} <tipo_de_marcador>
```

Por ejemplo, si se tuviera un CSV con información de dos basurales se generaría un archivo similar al siguiente:

```
34.902503,-57.9381686 {microbasural: Hipódromo de La Plata} <yellow-dot>
34.903492,-57.9376769 {punto de arroyo: Vías del tren} <tan-dot>
```

Dependiendo de la tipología del basural se debe identificar al sitio con un marcador de un color específico, como se indica a continuación:

```
punto de arroj o <tan-dot>
microbasural <yellow-dot>
basural <orange-dot>
macrobasural <default-dot>
```

Parte 2

Entrega: Semana 15/07/2019

Esta funcionalidad se ejecutará si el usuario ingresa los argumentos `-l`, `-s` y `-m` (o sus versiones largas):

Asignar un valor numérico a cada tipología de basural:

- punto de arroj o: 1
- microbasural: 2
- basural: 3
- macrobasural: 4

Para cada municipio (discriminado por `municipios_id`) sumar los valores de sus basurales. Y generar un reporte para `mapcustomizer.com` que tenga una sola entrada por cada municipio siguiendo las reglas indicadas a continuación:

- Por cada municipio se sumarán los valores de sus basurales según el listado anterior.
- Se tomará el municipio cuya suma de valores de basurales sea la máxima.
- Llamaremos `GRADO_RELATIVO` a este valor máximo dividido por 4.
- Usando este valor se asignaran los siguientes colores de marcador a cada municipio:

```
Si el valor es menor a GRADO_RELATIVO <tan-dot>
Si el valor es menor a 2 * GRADO_RELATIVO <yellow-dot>
Si el valor es menor a 3 * GRADO_RELATIVO <orange-dot>
Si el valor es mayor <default-dot>
```

- Se generará único punto por cada municipio (usando las coordenadas de un basural cualquiera de ese municipio) usando el marcador calculado.

1. Modalidad de la entrega

El trabajo es individual y se debe subir por <http://catedras.info.unlp.edu.ar> antes de las 23:55 del día de entrega establecido.

Todo lo entregado se someterá a un testeo de similitudes usando el similarity tester de Dick Grune¹ seguido de una inspección visual de los casos sospechosos para detectar posibles copias, cualquier copia detectada es motivo suficiente para que los involucrados desapruében la cursada.

La entrega del 24/06/2019 **es obligatoria** (aunque entreguen incompleto o en blanco) y no tiene calificación, simplemente recibirán una retroalimentación para seguir avanzando.

La entrega del 15/07/2019 **es obligatoria** se calificará y debe incluir toda la funcionalidad pedida para aprobar. En caso de hacer una entrega incompleta en esta etapa o desaprobado recibirán una retroalimentación para poder completar y corregir el trabajo para la reentrega del 5/08/2019.

La entrega del 5/08/2019 es la entrega final, en este punto el trabajo deberá estar completo y ser correcto para aprobar. La cursada depende de la aprobación de este trabajo (además de los parcialitos).

1.1. Entrega

Implementación de la funcionalidad completa de la aplicación como se detalla más abajo.

La entrega se considerará aprobada si se cumplen las siguientes condiciones:

- Se realiza la entrega en tiempo y forma.
- El programa funciona como es especificado en este documento sin fallar con ningún archivo de datos ni combinación de parámetros.
- El programa permite cargar archivos con líneas arbitrariamente largas sin fallar (aunque puede fallar con el mensaje “Campo muy largo” si algún campo del CSV es inusualmente largo).
- El alumno asiste al coloquio y demuestra conocimiento sobre el código desarrollado.
- Se libera toda la memoria alguna vez alocada (aún si el programa informa un error).
- Se lee correctamente desde archivos.
- Se hace un manejo correcto de arreglos, estructuras de datos dinámicas y punteros.
- El programa está correctamente modularizado y usando correctamente los `.h`.
- El programa es legible por cualquier desarrollador y compila sin warnings.

¹http://dickgrune.com/Programs/similarity_tester/

Nota 1: Se puede usar cualquier funcionalidad incluida en GNU libe.

Dataset: https://datos.gob.ar/dataset/acumar-linea-base-basurales-2017-cuenca-matanza-riachuelo/archivo/acumar_1642723b-2ab6-42ad-9ea1-1985811f8d7e

Map customizer: <https://www.mapcustomizer.com>