

Detección de las unidades de variables dado un .csv de entrada.

Descripción del problema

Se tiene un archivo .csv con datos de variables relacionadas al negocio geo-minero-metalúrgico-financiero. Se necesita un algoritmo que infiera el tipo de dato de cada columna del csv, asumiendo que cada una contiene información relativa a una variable.

Una variable puede ser del tipo categórica o continua. Si es categórica su unidad es un texto (""), en cambio si es continua su unidad puede ser: un porcentaje (%), partes por millón (ppm), partes por billón (ppb) o gramos por tonelada (g/t).

Solución

- Input
 - NombreEjecutable.exe M1.csv, , M2.csv
- Output
 - Un json con el siguiente formato [1].

```
{"SW":{"M1.csv":{"headers":["xcentre","ycentre","zcentre","cut_gc","zmin_gc","oretype_gc","cus_gc","mot_gc"],"types":["Continuous","Continuous","Continuous","Continuous","Categorical","Categorical","Continuous","Continuous"],"units":["ppm","ppm","ppm","%","","","%","%"]},"M2.csv":{"headers":["xcentre","ycentre","zcentre","cut_gc","zmin_gc","oretype_gc","cus_gc","mot_gc"],"types":["Continuous","Continuous","Continuous","Continuous","Categorical","Categorical","Continuous","Continuous"],"units":["ppm","ppm","ppm","%","","","%","%"]}}}
```

Representación numérica

Descripción del problema

Se tiene un archivo .csv con datos de variables relacionadas al negocio geo-minero-metalúrgico-financiero. Se necesita un algoritmo que infiera (o definida por parámetro) la cantidad óptima visible de dígitos a mostrar para cada columna.

Solución

- Input
 - NombreEjecutable.exe M1.csv, , M2.csv
 - NombreEjecutable.exe M1.csv, , M2.csv number_decimals (optional)
- Output
 - Un json con el siguiente formato [1].
 - Los cambios deben verse reflejados en archivos copias a partir de los de entrada.

```
{
  "SW": {
    "M1.csv": {
      "headers": ["xcentre", "ycentre", "zcentre", "cut_gc", "zmin_gc", "oretype_gc", "cus_gc", "mot_gc"],
      "numbers": ["3", "3", "3", "3", "2", "2", "2", "3"]
    },
    "M2.csv": {
      "headers": ["xcentre", "ycentre", "zcentre", "cut_gc", "zmin_gc", "oretype_gc", "cus_gc", "mot_gc"],
      "numbers": ["3", "3", "3", "3", "2", "2", "2", "3"]
    }
  }
}
```

Generación de nombre de variables abreviados.

Descripción del problema

Se tiene un archivo .csv con datos de variables relacionadas al negocio geo-minero-metalúrgico-financiero. Se necesita un algoritmo que permita inferir (o definida por parámetro que se aplicaría a todo el conjunto) un nuevo nombre para cada variable a través de su abreviación. Cada abreviación no debe repetirse y deben estar lo más disjuncto posible entre ellos.

Solución

- Input
 - NombreEjecutable.exe ,M1.csv, , M2.csv
 - NombreEjecutable.exe ,M1.csv, , M2.csv number_letters (optional)
- Solución
 - Un json con el siguiente formato [1].

```
{
  "SW": {
    "M1.csv": {
      "headers": ["xcentre", "ycentre", "zcentre", "cut_gc", "zmin_gc", "oretype_gc", "cus_gc", "mot_gc"],
      "abbrev": ["x", "y", "z", "cut", "zmin", "ore", "cus", "mot"]
    },
    "M2.csv": {
      "headers": ["xcentre", "ycentre", "zcentre", "cut_gc", "zmin_gc", "oretype_gc", "oretype", "zmin"],
      "abbrev": ["x", "y", "z", "cut", "zmi_gc", "ore_gc", "ore", "zmi"]
    }
  }
}
```

Sobre los entregables:

- Las implementaciones de las soluciones deben estar realizadas en C++11 o superior compiladas en Windows con optimización.
- Se debe generar solo un .exe que contenga todas las implementaciones mencionadas anteriormente. Si es necesario, se pueden agregar “flags” extras al command line arguments [2] para diferenciar cada solución. Ejemplo: --action
- Las librerías en las referencias son solo recomendaciones.
- **Usar librerías que se puedan comercializar.**

- Para cada implementación el tiempo de cómputo no debe superar los 3 segundos. No considerar el tiempo de generar los archivos copias con la nueva representación numérica (igual debe ser rápido).
- Para cada entregable pueden ingresar archivos extras de entrada que alimenten el algoritmo con estadísticas o casos bases generales.
- Es ideal que en esta parte del desarrollo entrevisten a sus colegas mineros y geólogos.

[1]: <https://github.com/nlohmann/json>

[2]: <http://tclap.sourceforge.net/>