

Prueba práctica para Profesional de Sistemas e Informática

Motivación

En la empresa existe un proceso en el cual se carga un archivo de datos a una aplicación y los persiste en un sistema de almacenamiento "X" y posteriormente si se desea, se puede volver a regenerar este archivo a partir de los datos persistidos.

Por otro lado, un usuario con permisos especiales en la aplicación puede realizar modificaciones sobre los datos cargados, lo cual hace que al generar el archivo nuevamente desde la aplicación tenga diferencias con respecto al que originalmente se cargó.

En ocasiones se requiere determinar si entre el archivo original y el que se genera a través de la aplicación presentan diferencias en los datos que contienen (en qué línea y en qué dato de esta), la primera opción para dar solución es simplemente recorrer ambos archivos e ir comparando el mismo número de línea de cada archivo y si no son iguales se guarda en una lista el número de la línea, esta lista al final contiene todas las líneas en las que hay diferencia, este enfoque funciona siempre y cuando los datos estén ordenados o se puedan ordenar por algún tipo de identificador único en cada archivo y además que sea equivalente en los dos archivos.

El problema es que en estos archivos los datos no tienen un identificador único que permita ordénalos de forma inequívoca, es decir, que permita que en ambos archivos las líneas de datos queden en el mismo orden sin importar que alguno de los datos contenidos en la línea haya variado, además que pueden llegar a tener una cantidad diferente de líneas.

Por lo anterior, se debe tener un enfoque diferente para dar solución a lo requerido, para esto vamos a suponer que los archivos tienen la misma cantidad de líneas (ya que simplemente en el programa se pueden agregar la cantidad de líneas faltantes al archivo que se requiera con información por defecto y que las identifique como que fueron agregadas), luego lo que se debe determinar cuáles son las diferentes agrupaciones de comparaciones que se pueden hacer entre las líneas de ambos archivos, después se debe ponderar cada grupo de comparaciones, esta ponderación se debe hacer de tal forma que un valor igual 0 indique que no hay diferencias y entre más lejano de 0 indique que hay muchas diferencias, es decir que a través de la ponderación se pueda determinar cuál grupo de comparaciones es la que genera la menor diferencia y la agrupación que tenga esta menor diferencia es la que se tomara como válida para mostrarle al usuario las diferencias entre ambos archivos si las hay.

Sede Arauca

Calle 16 No. 20 – 56 Oficina 305, Ed. Arauca Centro
Arauca, Colombia | Tel: (+57 7) 8852933

E-mail: contacto@arios-ing.com

Sede Neiva

Carrera 9 No.10 – 34, Local C
Neiva, Colombia | Cel: (+57 7) 3165318940

www.arios-ing.com

De este segundo enfoque uno de los puntos importante es lo que esta resaltado en color gris y es lo que se espera que pueda resolver.

Para tenerlo más claro lo que se necesita entonces es determinar cuáles son todos los grupos de parejas no repetidas con las comparaciones posibles y válidas que se pueden presentar entre las líneas de los dos archivos.

Ejemplo 1:

Archivo 1

0
1

Archivo 2

0
1

- Las comparaciones posibles son (0, 0), (0, 1), (1, 0) y (1, 1)
- Los grupos de comparaciones serian 2:
 - [(0, 0), (1, 1)]
 - [(0, 1), (1, 0)]

Ejemplo 2:

Archivo 1

0
1
2

Archivo 2

0
1
2

- Las comparaciones posibles son (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1) y (2, 2)
- Los grupos de comparaciones serian 6:
 - [(0, 0), (1, 1), (2, 2)]
 - [(0, 0), (1, 2), (2, 1)]
 - [(0, 1), (1, 0), (2, 2)]
 - [(0, 1), (1, 2), (2, 0)]
 - [(0, 2), (1, 0), (2, 1)]
 - [(0, 2), (1, 1), (2, 0)]
 - [(1, 2), (0, 0), (2, 1)] No válida pues contiene las mismas comparaciones del grupo de la segunda fila.

Sede Arauca

Calle 16 No. 20 – 56 Oficina 305, Ed. Arauca Centro
Arauca, Colombia | Tel: (+57 7) 8852933

E-mail: contacto@arios-ing.com

Sede Neiva

Carrera 9 No.10 – 34, Local C
Neiva, Colombia | Cel: (+57 7) 3165318940

www.arios-ing.com

Nota: En los ejemplos las comparaciones se representan como (x, y) donde “x” es el número de la línea del archivo 1 y “y” es el número de la línea del archivo 2, es decir, (0, 1) representa la comparación entre la línea 0 de archivo 1 y la línea 1 del archivo 2.

Puntos para desarrollar.

1. Con base a lo anterior se debe desarrollar un programa (Python o Java) que reciba la cantidad de líneas que tienen los dos archivos y que como salida me arroje la cantidad de grupos de comparaciones posibles, válidos y únicos. También se debe mostrar cómo está compuesto cada uno de estos grupos.

Es importante tener en cuenta que estos grupos deben quedar almacenados en alguna variable (el tipo lo define Ud.), ya que se esta se utilizará como insumo para realizar las comparaciones que permitirán ponderar cada agrupación, valga la pena aclarar que esto último no hace parte de la prueba.

2. El programa deberá tener control de versiones con Git y este deberá ser cargado en el repositorio de GitLab, al cual se le dará acceso.
3. Incluir en el repositorio un archivo en Word con el nombre “Readme” donde se especifique los requerimientos y pasos para ejecutar el programa.
4. Dentro del mismo repositorio de versiones incluir un archivo en Word en donde describa que pasos siguió para desarrollar el programa, es decir, que consideraciones tuvo en cuenta, si siguió alguna metodología, utilizó alguna fórmula o librería especial, etc. También debe comentar que dificultades tuvo y como las sorteó.
5. En otro archivo en Word quiero exponer que otro enfoque se puede adoptar para dar solución a la comparación de los archivos teniendo en cuenta las consideraciones dadas en la descripción. Este archivo también debe quedar en el repositorio de versiones.

Para la prueba tienen 3 días calendario para resolverla y al finalizar deben enviar un correo a asesorsistemas@arios-ing.com con asunto “Fin prueba sistemas” e indicando que en el repositorio asignado ya se encuentra todo cargado.

Sede Arauca

Calle 16 No. 20 – 56 Oficina 305, Ed. Arauca Centro
Arauca, Colombia | Tel: (+57 7) 8852933

E-mail: contacto@arios-ing.com

Sede Neiva

Carrera 9 No.10 – 34, Local C
Neiva, Colombia | Cel: (+57 7) 3165318940

www.arios-ing.com