

# Tarea Individual #1 - 2s2021

## Ingeniería de Software



### 1. Descripción

Dos redes sociales populares han anunciado que trabajarán en conjunto para unir sus servicios en una única plataforma. Para realizar esta tarea se requiere combinar la información que guarda cada base de datos sobre las cuentas de usuarios registrados en cada plataforma, la cual ha sido exportada a archivos CSV con la información requerida. Otro requerimiento importante es analizar los datos combinados y recopilar cierta información de utilidad sobre los usuarios de la plataforma, para enviarles correos y notificaciones posteriormente.

### 2. Consideraciones importantes:

- Los datasets entregados contienen información sobre 30.000 usuarios registrados en ambas plataformas. Se asume que cada persona usó el mismo correo para registrarse en ambas plataformas.
- Se considera como inactivo a cualquier usuario registrado en la plataforma que no ha iniciado sesión hace 2 o más años. Considere la fecha actual como 30 de agosto de 2021.
- Ambas plataformas tienen la funcionalidad de “seguir” a otros usuarios. Cuando un usuario sigue a otro, significa que se suscribe a los contenidos publicados por él para que aparezcan en su página de inicio.
- La operación contraria a “seguir” es “tener seguidores”. La base de datos almacena las personas a las que sigue cada usuario, pero no los seguidores, con el fin de evitar redundancia.

### 3. Insumos:

Se adjuntan las siguientes fuentes de datos en formato CSV para ser procesados, combinados y representados en su modelo con sus respectivas relaciones.

Archivo	Descripción
dataset1.csv	<p>Contiene información sobre los usuarios registrados en la 1ra red social.</p> <ul style="list-style-type: none"> <li>• <b>id:</b> identificador del usuario.</li> <li>• <b>correo:</b> dirección de correo electrónico del usuario.</li> <li>• <b>ultima_conexión:</b> fecha de última conexión del usuario, en formato AAAA-MM-DD.</li> <li>• <b>siguiendo:</b> corresponde a todos los identificadores de los usuarios a los que sigue este usuario actualmente en esta red social.</li> </ul>
dataset2.csv	<p>Contiene información sobre los usuarios registrados en la 2da red social.</p> <ul style="list-style-type: none"> <li>• <b>id:</b> identificador del usuario.</li> <li>• <b>correo:</b> dirección de correo electrónico del usuario.</li> <li>• <b>ultima_conexión:</b> fecha de última conexión del usuario, en formato AAAA-MM-DD.</li> <li>• <b>siguiendo:</b> corresponde a todos los identificadores de los usuarios a los que sigue este usuario actualmente en esta red social.</li> </ul>

#### 4. Desafíos

El software debe leer como entrada la carpeta con todos los archivos csv que se adjuntan y como salida debe generar dos archivos: uno llamado **datos.csv**, que contiene la información combinada entre ambos datasets y otro llamado **resultados.txt**, que contiene un breve reporte con los resultados.

El archivo **datos.csv** debe combinar la información de ambos datasets siguiendo las siguientes reglas:

1. Los IDs y correos se mantienen.
2. Se conserva la fecha de última conexión más reciente.
3. Combinar los usuarios seguidos por cada cuenta y ordenar estos IDs de menor a mayor.
4. El archivo resultante debe quedar con el mismo formato que los datasets originales, incluyendo el encabezado que indica los nombres de cada columna.

El archivo **resultados.txt** debe partir con su número de matrícula para identificación. El formato del archivo con lo que debe entregar se describe a continuación:

##### 1. Todos los usuarios inactivos.

```
### MATRÍCULA ###  
  
### INICIO USUARIOS INACTIVOS ###  
  
id;correo;ultima_conexion  
etc  
  
### FIN USUARIOS INACTIVOS ###
```

##### 2. Todos los usuarios que, entre las cuentas a las que sigue, al menos la mitad de ellas están inactivas.

```
### INICIO USUARIOS SIGUIENDO 50% O MÁS INACTIVOS ###  
  
id;correo;id_inactivo_1,...,id_inactivo_n  
etc  
  
### FIN USUARIOS SIGUIENDO 50% O MÁS INACTIVOS ###
```

##### 3. El usuario con la mayor cantidad de seguidores en total (o los usuarios, en caso de empate).

```
### INICIO USUARIOS CON MÁS SEGUIDORES ###  
  
id;correo;nro_seguidores  
etc  
  
### FIN USUARIOS CON MÁS SEGUIDORES ###
```

El símbolo # indicará al revisor cuál es el inicio y el término de dicha información por lo que es obligación que se incluyan.

## 5. Ejecución:

- Debe usar la plantilla de proyecto que se adjunta, no otra.
- La ejecución del programa debe ser por línea de comandos, solo a través del uso del ejecutable JAR, sin necesidad de tener que usar el IDE para ejecutar el programa.
- Se debe pasar por líneas de comandos como parámetros la carpeta con los archivos CSV (datasets), el nombre del archivo CSV combinado “datos.csv” y el nombre del archivo con los análisis: “resultados.txt”.
- A modo de ejemplo se adjunta la forma en que debería ejecutarse la aplicación:

```
java -jar mi_app.jar datasets datos.csv resultados.txt
```

Lo anterior debería tener la siguiente estructura de directorio:

```
├── mi_app.jar
├── datos.csv
├── resultados.txt
├── datasets
│   ├── dataset1.csv
│   └── dataset2.csv
```

- El archivo con las salidas debe ser reescrito cada vez que se ejecute el archivo jar.

## 6. Consideraciones de la entrega:

- Debe entregar en un archivo **zip**, con los siguientes elementos:
  - El código fuente con su proyecto en formato Maven.
  - El diagrama de clases (**imagen**) del modelo donde se expliciten las relaciones entre las clases que modelan el problema a partir de las fuentes de datos entregadas.
  - Considere que cada uno de los desafíos debe contener al menos una prueba unitaria.
  - El archivo jar ejecutable.
  - NO DEBE INCLUIR Los archivos CSV del dataset.
- **Tenga presente que el plagio parcial o total es una falta grave, por lo que es calificado con nota mínima e informado al director de carrera, arriesgando sanciones académicas.**