

# Badge 1

Serafí Nebot Ginard

# File Creator 1 - CSV -> XML

Procesa un fichero CSV y conviértelo a XML.

- Aplicación de consola (args).
- Generar el XML a través de nodos y estructura de árbol.
- Utiliza librerías que faciliten la tarea.
- ¿Puedes procesar cualquier CSV?

# Planteamiento

Un fichero CSV está formado por filas, separadas por un retorno de carro, y columnas, separadas por una ‘,’.

ID	Nombre	Apellidos	Edad
1	Serafi	Nebot, Ginard	19
2	Pepito	Palotes	32

Leemos el fichero por líneas. Cada línea la separamos por columnas y obtenemos una celda. Cada celda contiene un valor.

# ~~Problemas~~ Retos

Puede haber diferentes separadores dependiendo del fichero.

La primera fila puede contener el nombre de las columnas o valores.

Puede que haya un valor que contenga un separador.

El fichero puede contener caracteres inválidos en XML.

# Separadores

Hay diferentes tipos de separadores en un fichero CSV.

- Líneas (retorno de carro).
- Valores (,).
- Texto ("").

```
ID,Nombre,Apellidos,Edad  
1,Serafí,"Nebot, Ginard",19  
2,Pepito,Palotes,32
```

# Solución

Al ser una aplicación por consola le puedo indicar diferentes tipos de argumentos.

```
usage: file-creator-1
  --first-line-name      indica si la primera línea del archivo son
                        las columnas
  -i,--input <arg>      archivo de entrada
  -l,--line-separator <arg> separador de líneas ('\n' por defecto)
  --lowercase-tags       genera los tags del xml en minúscula
  -o,--output <arg>      archivo de salida
  -s,--separator <arg>  separador de valores (',' por defecto)
  -t,--text-separator <arg> separador de texto ('"' por defecto)
Error al procesar los argumentos
```

Librería: [org.apache.commons.cli](http://org.apache.commons.cli)

# ~~Problemas Retos~~

~~Puede haber diferentes separadores dependiendo del fichero.~~

~~La primera fila puede contener el nombre de las columnas o valores.~~

~~Puede que haya un valor que contenga un separador.~~

~~El fichero puede contener caracteres inválidos en XML.~~

# Caracteres inválidos XML

Hay una serie de caracteres que son inválidos en XML. Algunos se pueden escapar:

- < - &lt;
- > - &gt;
- ' - &apos;
- " - &quot;
- & - &amp;

He utilizado una librería para validar caracteres:

[org.apache.commons.text](http://org.apache.commons.text)

Contiene la utilidad `StringEscapeUtils` que permite escapar caracteres para diferentes formatos de ficheros.



# Caracteres inválidos XML

```
Identifier/Code,Access code;,Recovery\ code,"First name",'Last name',<Department>,&
Location
```

```
9012/1234,<12se74>,\rb9012;,"Rachel, & Mike",'Booker',Sales,Manchester
```

```
-<element>
  <Location>Manchester</Location>
  <Department>Sales</Department>
  <Access-code>&lt;12se74&gt;</Access-code>
  <First-name>Rachel, & Mike</First-name>
  <Last-name>&apos;Booker&apos;</Last-name>
  <IdentifierCode>9012/1234</IdentifierCode>
  <Recovery-code>\rb9012;</Recovery-code>
</element>
```