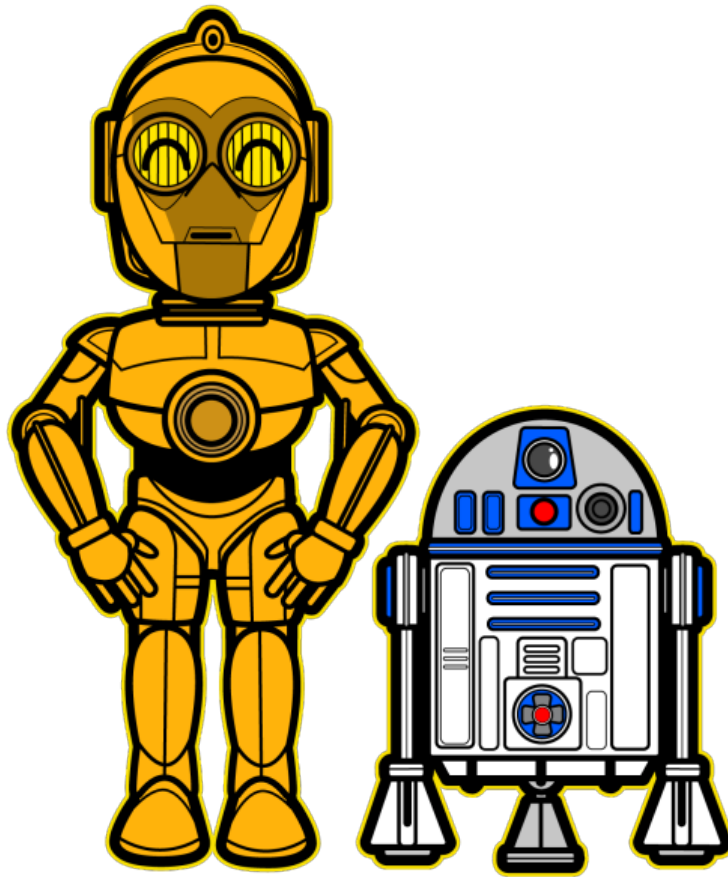


TP4 Automatas y gramaticas:

*Seguimiento de los usuarios que se han conectado días
feriados y no laborables*



Alumnos:

- Marotta, Alejandro Adrián
- Soria Gava, Lucas Damián

Consigna:

Seguimiento de los usuarios que se han conectado días feriados y no laborables

Solución:

El programa realiza un análisis del archivo .txt, que incluye registros de tráfico de conexiones de wifi, leyendo línea por línea y analizando si coincide con la expresión regular que desarrollaremos más adelante. Si no coincide con la misma, se descarta y se contabiliza como una línea incorrecta, de lo contrario se procede a realizar el análisis de los datos.

En el archivo las líneas están conformadas por los siguientes campos separados por punto y coma:

- ID Conexión unico
- Usuario
- Inicio de Conexión
- Fin de Conexio
- Session Time
- Input Octets
- Output Octets
- MAC AP
- MAC Cliente

Expresión regular:

La expresión regular creada para analizar los datos es la siguiente:

```
'^(?P<idcon>[0-9a-z]{16});(?P<usuario>(.[+]);(?P<fechaini>(([0-2][0-9])|(3[01]))V(0[1-9]|1[0-2])\d{4} ((1?[0-9])|(2[0-3])):([0-5][0-9]));(?P<fechafin>(([0-2][0-9])|(3[01]))V(0[1-9]|1[0-2])\d{4} ((1?[0-9])|(2[0-3])):([0-5][0-9]));(?P<tiempo>(\d+));(?P<input>(\d+));(?P<output>(\d+));(?P<macap>([A-F0-9]{2}-){5}[A-F0-9]{2}:UM));(?P<maccliente>([A-F0-9]{2}-){5}[A-F0-9]{2}))$'
```

Esta regex está dividida en grupos para facilitar el manejo de la lógica del programa, con esto podemos trabajar con cada uno de los datos sin necesidad de manipular Strings posteriormente.

idcon:

```
(?P<idcon>[0-9a-z]{16})
```

El Id de Conexión único está conformado por 16 caracteres alfanuméricos, en los cuales las letras están solo en minúscula.

Con esto encontramos un error en el archivo, el cual se hace presente cuando el id es solo de números, el sistema lo detecta como un número y lo coloca en notación científica: 8,37E+15. De esta forma, se pierden decimales y el id puede repetirse, haciéndolo inútil para el análisis.

usuario:

(?P<usuario>(.*))

Como no sabemos las reglas de la empresa para la creación de los nombres de usuarios, decidimos colocar que este pueda ser cualquier carácter al menos una vez.

fechaini y fechafin:

(?P<fechaini>(((0-2)[0-9]))(3[01]))\V(0[1-9]|1[0-2])\Vd{4} ((1?[0-9])|(2[0-3])):([0-5][0-9]))
(?P<fechafin>(((0-2)[0-9]))(3[01]))\V(0[1-9]|1[0-2])\Vd{4} ((1?[0-9])|(2[0-3])):([0-5][0-9]))

En este caso, ambas fechas se representan de la misma forma, pero decidimos hacer una distinción entre ambas, porque corresponden a dos datos distintos. Si lo hubiésemos tomado como:

((([0-2][0-9]))(3[01]))\V(0[1-9]|1[0-2])\Vd{4} ((1?[0-9])|(2[0-3])):([0-5][0-9]);}{2}

No nos habría permitido distinguir entre ambas fechas.

El formato que verifica son, por ejemplo, los siguientes:

23/09/2019 9:58

23/09/2019 11:11

tiempo:

(?P<tiempo>(\d+))

Para el tiempo de conexión decidimos tomar cualquier cantidad de dígitos concatenados, pero debe haber por lo menos uno.

input y output:

(?P<input>(\d+))

(?P<output>(\d+))

Cumplen con la misma lógica que el tiempo.

macap y maccliente:

```
(?P<macap>(([A-F0-9]{2}-){5}[A-F0-9]{2}:UM))  
(?P<maccliente>(([A-F0-9]{2}-){5}[A-F0-9]{2}))
```

En las redes de computadoras, la dirección MAC es un identificador de 48 bits (6 bloques de dos caracteres hexadecimales [8 bits]) que corresponde de forma única a una tarjeta o dispositivo de red. Se la conoce también como dirección física, y es única para cada dispositivo.

Formato:

B4-D1-DA-C4-45-42

Cabe destacar que en el archivo con los registros, las mac poseen sólo letras mayúsculas.

Las Mac de los Access Point, a diferencia de las MAC de los clientes, terminan con un “:UM”, ej:
04-18-D6-22-9B-4C:UM

Funcionamiento:

El programa está dividido en tres partes principales, las cuales antes de ser ejecutadas le preguntan al usuario si quieren visualizar esa información en particular.

La primera analiza los datos y toma todos los nombres de usuario y la lista de MACs, tanto de APs, como de clientes.

La segunda muestra todas las fechas de feriados y fines de semana, el día que representa esa fecha (sábado, domingo o el nombre de la fecha festiva) y la lista de usuarios que se conectaron a la red. Además, al principio imprime un calendario dinámico que se adapta a las fechas presentes en el archivo para ubicarnos dentro de los meses en los que realizamos el análisis.

Por último, la tercera parte permite buscar los días en los que se conectó un usuario en particular.

Utilidad:

Este software permite conocer los usuarios que se conectaron a la red de la empresa en días no laborables y hacer un seguimiento de cada uno en particular.