

MEMORIA

ALF



Nombre Completo	Subgrupo	DNI
Cuenca Guardiola, Patricia	2.4	
Pedreño García, Juan	2.4	

Convocatoria: enero 2024/25

Profesor: Juan Antonio Sánchez Laguna

Índice

1. Descripción de la aplicación.....	2
2. Manual de usuario.....	2
3. Formato de los datos.....	3
4. Aspectos principales.....	4
5. Conclusiones.....	6

1. Descripción de la aplicación

El objetivo de esta aplicación es poder realizar un filtrado de un fichero que contiene una lista de compras, con ciertos parámetros: número de teléfono, NIF, fecha, coordenadas, nombre de producto y precio. El filtrado que se quiera realizar deberá especificarse ajustando los parámetros de entrada mediante los cuales se invoca a la aplicación.

Nuestra aplicación es capaz de realizar una normalización del fichero, mediante la opción “-n”, convirtiendo las fechas al primer formato (YYYY-MM-DD HH:MM) y las coordenadas a GPS. Asimismo, mediante la opción “-sphone” se pueden buscar las compras realizadas por usuario con el número de teléfono que se especifique como parámetro. Con “-snif”, se realiza un filtrado similar al explicado anteriormente, sin embargo, con este comando se mostrará las compras realizadas por el usuario con el NIF especificado. Por otro lado, el comando “-stime” muestra las compras realizadas en un intervalo especificado, primero se indica el desde y posteriormente el hasta. Por último, nuestra aplicación puede realizar un filtrado de las entradas que se encuentran a cierta distancia(hasta) de una coordenada origen(desde), primero se especifica el desde y después el hasta en km.

2. Manual de usuario

Los comandos explicados anteriormente funcionan de la siguiente forma:

- -n <fichero>

El comando -n realiza un filtrado del fichero (segundo argumento), normalizando la fecha y las coordenadas

- -sphone <teléfono> <fichero>

El comando -sphone filtra el fichero (último argumento) mostrando las entradas cuyo teléfono coincida con el segundo argumento

- -snif <NIF> <fichero>

El comando -snif filtra el fichero (último argumento) mostrando las entradas cuyo NIF coincida con el segundo argumento

- -stime <desde> <hasta> <fichero>

La opción -stime filtra el fichero (último argumento) mostrando las entradas cuya fecha se encuentre en el intervalo desde la fecha, pasada como segundo argumento, hasta la fecha pasada como tercer argumento.

- -slocation <desde> <hasta> <fichero>

La opción -slocation filtra el fichero (último argumento) mostrando las entradas cuyas coordenadas se encuentren dentro del intervalo desde la coordenada de origen, pasada como segundo argumento, hasta la distancia máxima en kilómetros, pasada como tercer argumento.

3. Formato de los datos

El formato empleado para los teléfonos comprende 9 dígitos agrupados de tres en tres con un espacio en blanco entre los grupos. Además, admite teléfonos precedidos con un '+' de una longitud variable entre 10 y 15 números, agrupados como se desee.

Por otro lado, el NIF se puede expresar tanto como DNI, como NIE, es decir, 8 dígitos seguidos de una de las 23 letras del alfabeto cuando se trate de un DNI, o las letras X, Y o Z seguidas de 7 dígitos y terminando en una de las 23 letras del alfabeto cuando se trate de un NIE.

Asimismo, las fechas se pueden expresar en tres formatos. En el formato 1, el año el mes y el día que se indican separados por guiones, todos los números deben tener ceros para rellenar la cantidad máxima de dígitos, tras unos espacios, la hora y los minutos separados por el símbolo de los dos puntos, rellenando con ceros cuando sea necesario. El formato 2, aparece el mes expresado mediante su nombre en inglés y, separado por espacios, el día del mes. Seguidamente, separado por una coma y tras uno o más espacios, aparece la hora y los minutos separados por dos puntos en el sistema horario de 12 horas y seguidamente las letras am y pm, tanto en mayúsculas como en minúsculas. Por último, el tercer formato, aparece la hora expresada mediante tres grupos de dos dígitos (rellenando con cero cuando sea necesario) separados por el símbolo de los dos puntos, en este caso, aparecen los segundos. Tras uno o más espacios aparecen tres grupos de dígitos separados por el símbolo de la barra de dividir. Primero el día, después el mes y el año. En este caso también se rellenan con cero.

Las coordenadas tienen 3 formatos, el primero es expresando los ángulos en decimal. Primero la latitud y después la longitud. Estos están separados por una coma. Y pueden tener un número indefinido de decimales. Si uno de ellos es negativo será precedido por el signo "-" pero si es positivo puede ir o no ir precedido del signo "+". El segundo formato es el sexagesimal. Cada ordenada tiene el ángulo seguido por el símbolo "°", los minutos seguidos de una comilla y los segundos seguidos de dobles comillas. Después, cada una de ellas tiene una letra que representa en el caso de la latitud si es Norte o Sur y la longitud si es Este o Oeste.

Por último, para el producto no hay ningún tipo de restricciones. Sin embargo, el precio tiene que ser un número, entero o real, con el símbolo del €.

4. Aspectos principales

La expresión regular de teléfono o bien toma 9 dígitos divididos en grupos de 3 sin prefijo, o bien una serie de dígitos precedidos de un más agrupados de cualquier forma. Para asegurar que cumpla con la longitud establecida se encarga la lógica del código, no la expresión regular.

```
r"^\d{3} \d{3} \d{3}\s*$|+(\d+\s*)+"
```

La expresión regular de el NIF es: `r"(?P<num>\d{8}|[X-Z]\d{7})(?P<letra>[A-Z])"`

Esta expresión identifica dos grupos, el primero que son los dígitos(en el caso del NIF también incluye la primera letra), y la letra final.

La expresión regular del primer formato de fecha es:

```
r"(?P<anho>\d{4})-(?P<mes>[0-1]\d)-(?P<dia>[0-3]\d)\s+(?P<hora>[0-2]\d):  
(?P<minutos>[0-5]\d)(?P<segundos>)(?P<am>)"
```

Esta expresión identifica el año como los 4 primeros dígitos, después el mes, un número, desde uno (01) hasta doce, y los días que van de uno (01) a treinta y uno, separados por guiones y rellenando con ceros cuando sea necesario. A continuación, están las horas y los minutos separados por ":" y rellenando también cuando sea necesario. Los segundos no están presentes en este formato.

La expresión regular del segundo formato de fecha es:

```
r"(?i)(?P<mes>January|February|March|April|May|June|July|August|September|October|  
November|December)\s+(?P<dia>[1-3]\d|\d),\s+  
(?P<anho>\d|\d{2}|\d{3}|\d{4})\s+(?P<hora>[1]\d|\d):(?P<minutos>[0-5]*\d)\s+  
(?i)(?P<am>AM|PM)(?P<segundos>)"
```

Esta expresión identifica el mes escrito en inglés. Lo identifica en mayúsculas o minúsculas y cualquier combinación de las mismas. Tras uno o más espacios, aparece el día. Después de una coma y uno o más espacios, está el año que puede ir desde 0 hasta 9999. Seguidamente, y tras uno o más espacios, van la hora, en sistema horario de 12 horas; los minutos; uno o más espacios; y si es AM o PM. En este formato, solo hace falta rellenar con ceros los minutos y da igual si el am/pm está en mayúscula o minúscula.

La expresión regular de tercer formato de fecha es:

```
r" (?P<hora>[0-2]\d) : (?P<minutos>[0-5]\d) : (?P<segundos>[0-5]\d) \s+  
(?P<dia>[0-3]\d) / (?P<mes>[0-1]\d) / (?P<anho>\d{4}) (?P<am>) "
```

En este formato, primero identifica la hora en formato 24 horas seguido de ":" y los minutos y otros ":" y los segundos. Luego un espacio y los días seguido de "/" y el mes seguido de "/" y el año. También rellenando con ceros cuando sea necesario.

La expresión regular para las coordenadas decimales es:

```
(?P<latitud>[+,-]?[0-1]*\d+.\d+), \s+(?P<longitud>[+,-]?[0-1]*\d+.\d+)
```

En este formato, las coordenadas son un número positivo, que pueden llevar o no el símbolo '+' o un número negativo que debe llevar '-' con uno o más decimales. La latitud es el primer número y está separado de la longitud por una coma y uno o más espacios.

La expresión regular para las coordenadas sexagesimales:

```
(?P<latitudg>\d+) ° (?P<latitudmin>[0-5]*\d) ' (?P<latituds>[0-5]*\d.\d+)\ "  
(?P<n>N|S), \s*(?P<longitudg>1?\d*) ° (?P<longitudmin>[0-5]*\d) '  
(?P<longituds>[0-5]*\d.\d+)\ " (?P<e>E|W)
```

En este formato, los ángulos están expresados en formato sexagesimal, es decir, con grados, minutos y segundos, las coordenadas son positivas, y tras cada una hay una letra, para la latitud puede ser N o S y para la longitud E o W. La latitud y la longitud están separados por una coma y cero o más espacios.

La expresión regular para las coordenadas GPS es:

```
(?P<latitudg>[0-1]\d{2})(?P<latitudmin>[0-5]\d)(?P<latituds>[0-5]\d.\d{4})  
(?P<n>N|S)(?P<longitudg>[0-1]\d{2})(?P<longitudmin>[0-5])  
(?P<longituds>\d{0-5}\d.\d{4})(?P<e>E|W)
```

En este caso, se expresan las coordenadas en formato sexagesimal, sin incluir el símbolo de grados, ni minutos, ni segundos. Tampoco incluye comas ni espacios entre las coordenadas. Los grados ocupan tres dígitos, los minutos dos y los segundos 2 dígitos un punto '.' y cuatro decimales y una letra (N o S) para la latitud y otra letra (E o W) para la longitud.

La expresión regular para el precio es un número (o números) cualquiera que puede o no tener una coma seguida de más números y posteriormente el símbolo del euro:
`r"(?P<Precio>[\d]+(?:\.\d+)?) [€]"`

5. Conclusiones

Este proyecto nos ha resultado interesante en varios sentidos. El primero de ellos es que hemos usado un lenguaje que aún no habíamos usado en la carrera. Al comparar Python con otros lenguajes de programación usados para el desarrollo de otras asignaturas, que no facilitan tanto la programación, nos ha fascinado la comodidad a la hora de trabajar. Por otro lado, consideramos que los desafíos que hemos tenido que resolver nos han parecido útiles de cara a futuros problemas que nos podamos encontrar en otros proyectos. Un ejemplo concreto es el control de excepciones, pues hemos podido preparar una aplicación robusta frente a datos de entrada con errores de formato, algo que es esencial en un programa real.

Por otra parte, las mayores dificultades las hemos tenido con las expresiones regulares, las cuáles hemos tenido que revisar bastantes veces, conforme íbamos encontrando casos nuevos que hacían que fallasen para poder mejorarlas lo máximo posible, sin que fueran muy específicas ni muy generales. La manera de superar esos problemas era siendo más meticulosos con los casos de prueba que le pasábamos.

En definitiva, nuestro grado de satisfacción con la asignatura y el proyecto es alto, ya que mucha parte del trabajo la hemos podido realizar en las sesiones de prácticas consultándole al profesor las dudas que nos iban surgiendo. Además, consideramos que el proyecto ha sido muy beneficioso para mejorar nuestras habilidades de programación, ya que al usar Python, nos hemos podido centrar en la tarea que teníamos que realizar y no en cómo gestionar la memoria dinámica.