Programación Avanzada Daniel Gregorio Longino Tarea 10

Una de las tareas más comunes que puede realizar con Python es leer y escribir archivos. Ya sea escribiendo en un archivo de texto simple, leyendo un registro de servidor complicado o incluso analizando datos de bytes sin procesar, todas estas situaciones requieren leer o escribir un archivo.

1. ¿Qué es un archivo?

Antes de que podamos analizar cómo trabajar con archivos en Python, es importante comprender qué es exactamente un archivo y cómo los sistemas operativos modernos manejan algunos de sus aspectos.

En esencia, un archivo es un conjunto contiguo de bytes que se utiliza para almacenar datos. Estos datos están organizados en un formato específico y pueden ser algo tan simple como un archivo de texto o tan complicado como un programa ejecutable. Al final, estos archivos de bytes se traducen a 1 y 0 binarios para que la computadora los procese más fácilmente.

Los archivos en la mayoría de los sistemas de archivos modernos se componen de tres partes principales:

- 1. Encabezado. Metadatos sobre el contenido del archivo (nombre de archivo, tamaño, tipo, etc.)
- 2. Datos: contenido del archivo escrito por el creador o editor.
- 3. Final del archivo: carácter especial que indica el final del archivo

1.1. Atributos del objeto archivo

El objeto de archivo tiene los siguientes atributos:

- name: Devuelve el nombre del archivo. Es un atributo de solo lectura. Si el objeto de archivo se creó con la función open(), se devuelve el nombre del archivo. De lo contrario, se devuelve alguna cadena indica el origen del objeto de archivo.
- 2. encoding: devuelve la codificación que utiliza este archivo, como UTF-8. Este atributo es de solo lectura. Cuando las cadenas Unicode se escriben en un archivo, se convertirán en cadenas de bytes utilizando esta codificación. También puede ser None. En ese caso, el archivo usa la codificación predeterminada del sistema para convertir cadenas Unicode.

- 3. Devuelve un valor booleano que indica si se debe imprimir un carácter de espacio antes de otro valor cuando se usa la declaración de impresión.
 - mode: Devuelve el modo de acceso del archivo utilizado.
 - closed: Devuelve True si un archivo está cerrado. Es un valor booleano que indica el estado actual del objeto de archivo.
- 4. errors: es una cadena opcional que especifica cómo se deben manejar los errores de codificación y decodificación; esto no se puede usar en modo binario. Pase 'strict' para generar una excepción ValueError si hay un error de codificación (el valor predeterminado None tiene el mismo efecto), o pase 'ignore' para ignorar los errores.
- 5. newline: los archivos abiertos en el modo universal de lectura de nueva línea realizan un seguimiento de las nuevas líneas encontradas al leer el archivo. Los valores son '\r', '\n', '\r', '\n', None (todavía no se han leído nuevas líneas) o una tupla que contiene todos los tipos de nuevas líneas vistos. Para los archivos que no se abren en el modo de lectura universal de nueva línea, el valor de este atributo será None.

```
with open(r'E:\pynative\files\test.txt', "r") as fp:
    print('Is Closed:', fp.closed)
    print('Encoding Used:', fp.encoding)
    print('Access Mode:', fp.mode)
    print('NewLines Found:', fp.newlines)
```

1.2. Rutas de archivos

Cuando accede a un archivo en un sistema operativo, se requiere una ruta de archivo. La ruta del archivo es una cadena que representa la ubicación de un archivo. Se divide en tres partes principales:

- 1. Ruta de la carpeta: la ubicación de la carpeta de archivos en el sistema de archivos donde las carpetas subsiguientes están separadas por una barra diagonal / (Unix) o una barra diagonal inversa \ (Windows).
- 2. Nombre de archivo: el nombre real del archivo.
- 3. Extensión: el final de la ruta del archivo precedido por un punto (.) utilizado para indicar el tipo de archivo.

1.3. Finales de línea

Un problema que se encuentra a menudo cuando se trabaja con datos de archivos es la representación de una nueva línea o final de línea. El final de línea tiene sus raíces en la era del Código Morse, cuando se usaba un signo específico para comunicar el final de una transmisión o el final de una línea. Más tarde, esto fue estandarizado para teleimpresores tanto por la Organización Internacional de Normalización (OIN) como por la Asociación Estadounidense de Normalización (AEN). El estándar OIN establece que los finales de línea deben usar la secuencia de caracteres de retorno de carro (CR o \r) y avance de línea (LF o \n) (CR+LF o \r\ n). Sin embargo, el estándar (OIN) permitía los caracteres CR+LF o solo el carácter LF.

Windows usa los caracteres CR+LF para indicar una nueva línea, mientras que Unix y las versiones más nuevas de Mac usan solo el carácter LF. Esto puede causar algunas complicaciones cuando procesa archivos en un sistema operativo que es diferente al origen del archivo. Aquí hay un ejemplo rápido. Digamos que examinamos el archivo dog_breeds.txt que se creó en un sistema Windows:

Pug\r\n
Jack Russell Terrier\r\n
English Springer Spaniel\r\n
German Shepherd\r\n
Staffordshire Bull Terrier\r\n
Cavalier King Charles Spaniel\r\n
Golden Retriever\r\n
West Highland White Terrier\r\n
Boxer\r\n
Border Terrier\r\n

1.4. Codificaciones de caracteres

Otro problema común que puede enfrentar es la codificación de los datos de bytes. Una codificación es una traducción de datos de bytes a caracteres legibles por humanos. Esto normalmente se hace asignando un valor numérico para representar un carácter. Las dos codificaciones más comunes son los formatos ASCII y UNICODE. ASCII solo puede almacenar 128 caracteres, mientras que Unicode puede contener hasta 1,114,112 caracteres.

ASCII es en realidad un subconjunto de Unicode (UTF-8), lo que significa que ASCII y Unicode comparten los mismos valores numéricos a caracteres. Es importante tener en cuenta que analizar un archivo con la codificación de caracteres incorrecta puede provocar fallas o tergiversar el carácter. Por ejemplo, si un archivo se creó con la codificación UTF-8 e intenta analizarlo con la codificación ASCII, si hay un carácter que está fuera de esos 128 valores, se generará

un error.

2. Métodos del objeto file

El objeto File, entre sus métodos más frecuentes, dispone de los siguientes:

2.1. open() y close()

Cuando desee trabajar con un archivo, lo primero que debe hacer es abrirlo. Esto se hace invocando la función integrada open(). open() tiene un solo argumento requerido que es la ruta al archivo. open() tiene un solo retorno, el objeto de archivo:

```
Python

file = open('dog_breeds.txt')
```

Es importante recordar que es su responsabilidad cerrar el archivo. En la mayoría de los casos, al finalizar una aplicación o secuencia de comandos, el archivo se cerrará finalmente. Sin embargo, no hay garantía de cuándo sucederá exactamente eso. Esto puede conducir a un comportamiento no deseado, incluidas las fugas de recursos. También es una práctica recomendada dentro de Python asegurarse de que su código se comporte de una manera bien definida y reduzca cualquier comportamiento no deseado.

Cuando está manipulando un archivo, hay dos formas que puede usar para asegurarse de que un archivo se cierre correctamente, incluso cuando se produzca un error. La primera forma de cerrar un archivo es usar el bloque try-finally:

```
Python

reader = open('dog_breeds.txt')
try:
    # Further file processing goes here
finally:
    reader.close()
```

La segunda forma de cerrar un archivo es usar la instrucción with:

```
Python
with open('dog_breeds.txt') as reader:
    # Further file processing goes here
```

La instrucción with automáticamente se encarga de cerrar el archivo una vez que sale del bloque with, incluso en casos de error. Se recomienda encarecidamente que utilice la declaración with tanto como sea posible, ya que permite un código más limpio y facilita el manejo de cualquier error inesperado.

Lo más probable es que también desee utilizar el segundo argumento posicional, mode. Este argumento es una cadena que contiene varios caracteres para representar cómo desea abrir el archivo. El valor predeterminado y más común es 'r', que representa abrir el archivo en modo de solo lectura como un archivo de texto:

```
Python
with open('dog_breeds.txt', 'r') as reader:
    # Further file processing goes here
```

Otras opciones para mode son las siguientes:

- 1. 'r'. Abrir para lectura (predeterminado).
- 2. 'w'. Abrir para escribir.
- 3. 'rb' o 'wb'. Abrir en modo binario (lectura/escritura usando datos de byte)

Una vez que haya abierto un archivo, querrá leerlo o escribirlo. En primer lugar, cubramos la lectura de un archivo. Hay varios métodos a los que se puede llamar en un objeto de archivo para ayudarlo:

- .read(size=-1). size representa el número de bytes para leer de un archivo. Se devuelve una cadena vacía cuando se encuentra EOF inmediatamente. Si no se pasa ningún argumento o se pasa None o -1, entonces se lee todo el archivo.
- 2. .readline(size=-1). Si se proporciona size, lee el número de bytes (incluido el salto de línea final) de un archivo. Si el argumento de size es negativo o no se especifica, lee una sola línea. Se devuelve una cadena vacía cuando se encuentra un final de archivo inmediatamente.
- 3. .readlines(). Esto lee las líneas restantes del objeto de archivo y las devuelve como una lista.

Usando el mismo archivo dog_breeds.txt que usaste anteriormente, veamos algunos ejemplos de cómo usar estos métodos. Aquí hay un ejemplo de cómo abrir y leer el archivo completo usando .read():

Aquí hay un ejemplo de cómo leer 5 bytes de una línea cada vez usando el método Python .readline():

Aquí hay un ejemplo de cómo leer el archivo completo como una lista usando el método Python .readlines():

```
Python
>>> with open('dog_breeds.txt', 'r') as reader:
>>>
        # Read & print the entire file
        print(reader.read())
>>>
Pug
Jack Russell Terrier
English Springer Spaniel
German Shepherd
Staffordshire Bull Terrier
Cavalier King Charles Spaniel
Golden Retriever
West Highland White Terrier
Boxer
Border Terrier
```

```
Python
>>> with open('dog_breeds.txt', 'r') as reader:
        # Read & print the first 5 characters of the line 5 times
        print(reader.readline(5))
        # Notice that line is greater than the 5 chars and continues
>>>
        # down the line, reading 5 chars each time until the end of the
        # line and then "wraps" around
>>>
        print(reader.readline(5))
>>>
>>>
        print(reader.readline(5))
        print(reader.readline(5))
>>>
>>>
        print(reader.readline(5))
Pug
Jack
Russe
ll Te
rrier
```

2.2. write()

Ahora profundicemos en la escritura de archivos. Al igual que con la lectura de archivos, los objetos de archivo tienen varios métodos que son útiles para escribir en un archivo:

- 1. .write(string). Esto escribe la cadena string en el archivo.
- 2. .writelines(seq) Esto escribe la secuencia en el archivo. No se añaden finales de línea a cada elemento de secuencia. Depende de usted agregar los finales de línea apropiados.

```
with open(r'E:\pynative\files\test.txt', 'w') as fp:
    fp.write('My New Line')
```

```
with open(r'E:\pynative\files\test.txt', 'w') as fp:
   data = ['line1\n', 'line2\n', 'line3\n']
   fp.writelines(data)
```

2.3. truncate()

Utilice el método truncate() para dejar el archivo vacío. Si el argumento size está presente, el archivo se trunca a (como máximo) ese tamaño. Entonces, por ejemplo, si especifica 10 bytes, truncate() eliminará los primeros diez bytes de un archivo. El tamaño predeterminado es la posición actual de un puntero de archivo. La posición actual del archivo no cambia. Tenga en cuenta que si un tamaño especificado excede el tamaño actual del archivo, el resultado depende de la plataforma: las posibilidades incluyen que el archivo permanezca sin cambios, aumente al tamaño especificado como si estuviera lleno de ceros o aumente al tamaño especificado con contenido nuevo indefinido.

```
with open(r'E:\pynative\files\test.txt', 'a') as fp:
   fp.truncate(5)
```

2.4. seek() y tell()

La función seek() mueve el puntero hacia el byte indicado y la función tell() devuelve la posición actual del puntero.

```
with open(r'E:\pynative\files\test.txt', "r") as fp:
    # Moving the file handle to 6th character
    fp.seek(6)
    # read file
    print(fp.read())

# get current position of file pointer
    print(fp.tell())
```

2.5. fileno()

Devuelve el descriptor de archivo de enteros utilizado por el sistema de implementación subyacente para solicitar operaciones de I/O del sistema operativo. Esto puede ser útil para otras interfaces de nivel inferior que usan descriptores de archivos, como os.read().

```
with open(r'E:\pynative\files\test.txt', "r") as fp:
    print(fp.fileno())
# Output 3
```

2.6. flush()

Como sugiere el nombre, vacía el búfer interno. Cuando se utiliza el almacenamiento en búfer y si está escribiendo en un archivo, es posible que la línea no aparezca en el archivo hasta que se llame al método flush() o close().

```
with open(r'E:\pynative\files\test.txt', "w") as fp:
   fp.write('New Line')
   fp.flush()
```

2.7. issaty()

Devuelva True si el archivo está conectado a un dispositivo TTY como una teleimpresora, de lo contrario, False.

```
with open(r'E:\pynative\files\test.txt', "r") as fp:
    print(fp.isatty())
# Output False
```

2.8. readinto()

Lee hasta len(b) bytes en bytearray b y devuelva el número de bytes leídos. Si el objeto está en modo sin bloqueo y no hay bytes disponibles, se devuelve None.

2.9. next()

El método next() se usa cuando un archivo se usa como un iterador, generalmente en un bucle, el método next() se llama repetidamente. Este método devuelve la siguiente línea de entrada o genera StopIteration cuando se alcanza el final del archivo.

```
#!/usr/bin/python

# Open a file
fo = open("foo.txt", "rw+")
print "Name of the file: ", fo.name

# Assuming file has following 5 lines
# This is 1st line
# This is 2nd line
# This is 3rd line
# This is 4th line
# This is 5th line

for index in range(5):
    line = fo.next()
    print "Line No %d - %s" % (index, line)

# Close opend file
fo.close()
```

2.10. xreadlines()

Devuele un iterador sobre las lineas del archivo.

```
Name of the file: foo.txt
Line No 0 - This is 1st line
Line No 1 - This is 2nd line
Line No 2 - This is 3rd line
Line No 3 - This is 4th line
Line No 4 - This is 5th line
```

3. Pylint

Reporte generado por Pylint para las funciones date_in_spanish, from_stan-dard_equity_option_convention y collect_commission_adjustment.

```
laplacian spectra.aux
laplacian spectra.du
laplacian spectra.out
laplacian spectra.out
laplacian spectra.syncte
laplacian spectra.syncte
laplacian spectra.tex
main.py
matrices.py
matrices.py
matrices(1).py
progra.py
progra.py
progra.py
progra.py
progra.py
laplacian spectra.log
laplacian spectra.tex
main.py
laplacian spectra.tex
matrices(1).py
laplacian spectra.tex
laplacian spectra.te
```

```
Laplacian spectra.aux
laplacian spectra.out
laplacian spectra.out
laplacian spectra.ydf
```