



ESCUELA POLITÉCNICA NACIONAL

ANÁLISIS DE DATOS

TRABAJO EN CLASE



NOMBRE: Diana Almeida

FECHA: 10-06-2021

INSTRUCCIONES:

- El objetivo es contar el número de ocurrencias de las palabras en un archivo de texto. El resultado final deberá ser similar a:
(Harry,250)
(potter,153)
(niño,48)
- Subir en la plataforma el link del repositorio de github.
- En el repositorio se debe incluir al menos una captura de pantalla de la ejecución del programa.
- Código + dataset

DESARROLLO:

A continuación, se presenta el código utilizado para la creación del programa que permite contar el número de ocurrencias de las palabras en un archivo de texto.

```
print("<<  NUMERO DE OCURRENCIAS DE PALABRAS DE UN TEXTO  >>")
# Variable
opc=0
# Función para Leer Archivo txt:
def contarOcurrencias(i):
    if i == 1:
        try:
            with open("Julio Cortazar - Continuidad de los
parques.txt") as obj:
                contenido = obj.read()
                r= str(input("\n ¿Desea ver el contenido del libro?
(s/n) : "))
                if (r == 's'):
                    print("\n-----> INICIO DEL LIBRO\n")
                    print(contenido)
                    print("\n-----> FIN DEL LIBRO")
                except FileNotFoundError:
                    mensaje = ("No se puede abrir el archivo")
                    print(mensaje)
                else:
                    palabras = contenido.split()
                    numPalabras = len(palabras)
                    print("\nEl archivo contiene "+ str(numPalabras)+ "
palabras")
                    print("\n<<  OCURRENCIAS DE CADA PALABRA  >> \n")
                    ocurrencia = [palabras.count(p) for p in palabras]
                    print(str(list(zip(palabras, ocurrencia))))
        if i == 2:
            try:
                with open("Edgar Allan - El barril amontillado.txt") as
obj:
```

```

        contenido = obj.read()
        r = str(input("\n ¿Desea ver el contenido del libro?
(s/n) : "))

        if (r == 's'):
            print("\n-----> INICIO DEL LIBRO\n")
            print(contenido)
            print("\n-----> FIN DEL LIBRO")
        except FileNotFoundError:
            mensaje = ("No se puede abrir el archivo")
            print(mensaje)
        else:
            palabras = contenido.split()
            numPalabras = len(palabras)
            print("\nEl archivo contiene " + str(numPalabras) + "
palabras")

            print("\n<<  OCURRENCIAS DE CADA PALABRA  >> \n")
            ocurrencia = [palabras.count(p) for p in palabras]
            print(str(list(zip(palabras, ocurrencia))))

    if i == 3:
        try:
            with open("Oscar Wilde - El principe feliz.txt") as obj:
                contenido = obj.read()
                r = str(input("\n ¿Desea ver el contenido del libro?
(s/n) : "))

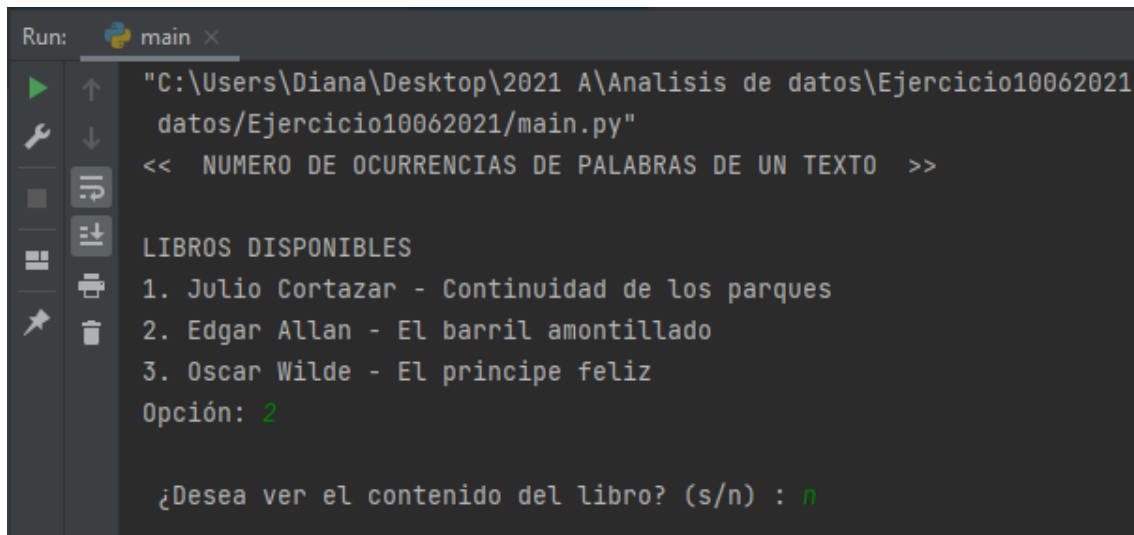
                if (r == 's'):
                    print("\n-----> INICIO DEL LIBRO\n")
                    print(contenido)
                    print("\n-----> FIN DEL LIBRO")
                except FileNotFoundError:
                    mensaje = ("No se puede abrir el archivo")
                    print(mensaje)
                else:
                    palabras = contenido.split()
                    numPalabras = len(palabras)
                    print("\nEl archivo contiene " + str(numPalabras) + "
palabras")

                    print("\n<<  OCURRENCIAS DE CADA PALABRA  >> \n")
                    ocurrencia = [palabras.count(p) for p in palabras]
                    print(str(list(zip(palabras, ocurrencia))))

# Menu
while opc < 1 or opc > 3:
    print("\nLIBROS DISPONIBLES")
    print("1. Julio Cortazar - Continuidad de los parques")
    print("2. Edgar Allan - El barril amontillado")
    print("3. Oscar Wilde - El principe feliz")
    opc = int(input("Opción: "))
    contarOcurrencias(opc)

```

CAPTURAS DE PANTALLA: Ejecución

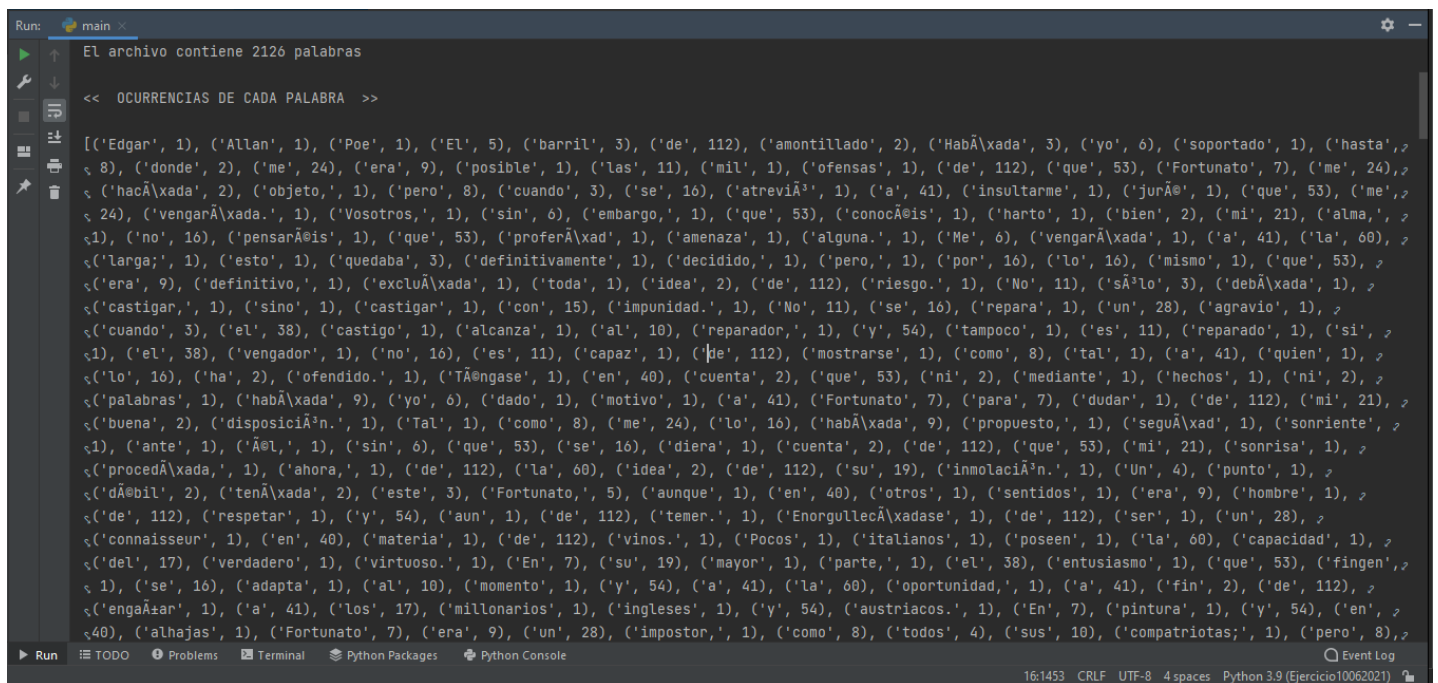


```
Run: main x
"C:\Users\Diana\Desktop\2021 A\Análisis de datos\Ejercicio10062021\
datos\Ejercicio10062021/main.py"
<< NUMERO DE OCURRENCIAS DE PALABRAS DE UN TEXTO >>

LIBROS DISPONIBLES
1. Julio Cortazar - Continuidad de los parques
2. Edgar Allan - El barril amontillado
3. Oscar Wilde - El principe feliz
Opción: 2

¿Desea ver el contenido del libro? (s/n) : n
```

Fig.1. Selección del usuario de 1 de 3 libros
y verificación de presentación de contenido del libro.



```
Run: main x
El archivo contiene 2126 palabras

<< OCURRENCIAS DE CADA PALABRA >>

[('Edgar', 1), ('Allan', 1), ('Poe', 1), ('El', 5), ('barril', 3), ('de', 112), ('amontillado', 2), ('Hab\xada', 3), ('yo', 6), ('soportado', 1), ('hasta', 8), ('donde', 2), ('me', 24), ('era', 9), ('posible', 1), ('las', 11), ('mil', 1), ('ofensas', 1), ('de', 112), ('que', 53), ('Fortunato', 7), ('me', 24), ('hac\xada', 2), ('objeto', 1), ('pero', 8), ('cuando', 3), ('se', 16), ('atrevi\xed', 1), ('a', 41), ('insultarme', 1), ('jur\xed', 1), ('que', 53), ('me', 24), ('vengar\xada', 1), ('Vosotros', 1), ('sin', 6), ('embargo', 1), ('que', 53), ('conoc\xedais', 1), ('harto', 1), ('bien', 2), ('mi', 21), ('alma', 1), ('no', 16), ('pensar\xedais', 1), ('que', 53), ('profer\xada', 1), ('amenaza', 1), ('alguna', 1), ('Me', 6), ('vengar\xada', 1), ('a', 41), ('la', 60), ('larga', 1), ('esto', 1), ('quedaba', 3), ('definitivamente', 1), ('decidido', 1), ('pero', 1), ('por', 16), ('lo', 16), ('mismo', 1), ('que', 53), ('era', 9), ('definitivo', 1), ('exclu\xada', 1), ('toda', 1), ('idea', 2), ('de', 112), ('riesgo', 1), ('No', 11), ('s\xed', 3), ('deb\xada', 1), ('castigar', 1), ('sino', 1), ('castigar', 1), ('con', 15), ('impunidad', 1), ('No', 11), ('se', 16), ('repara', 1), ('un', 28), ('agravio', 1), ('cuando', 3), ('el', 38), ('castigo', 1), ('alcanza', 1), ('al', 10), ('reparador', 1), ('y', 54), ('tampoco', 1), ('es', 11), ('reparado', 1), ('si', 1), ('el', 38), ('vengador', 1), ('no', 16), ('es', 11), ('capaz', 1), ('de', 112), ('mostrarse', 1), ('como', 8), ('tal', 1), ('a', 41), ('quien', 1), ('lo', 16), ('ha', 2), ('ofendido', 1), ('t\xedngase', 1), ('en', 40), ('cuenta', 2), ('que', 53), ('ni', 2), ('mediante', 1), ('hechos', 1), ('ni', 2), ('palabras', 1), ('hab\xada', 9), ('yo', 6), ('dado', 1), ('motivo', 1), ('a', 41), ('Fortunato', 7), ('para', 7), ('dudar', 1), ('de', 112), ('mi', 21), ('buena', 2), ('disposici\xed', 1), ('tal', 1), ('como', 8), ('me', 24), ('lo', 16), ('hab\xada', 9), ('propuesto', 1), ('segu\xada', 1), ('sonriente', 1), ('ante', 1), ('a', 41), ('sin', 6), ('que', 53), ('se', 16), ('diera', 1), ('cuenta', 2), ('de', 112), ('que', 53), ('mi', 21), ('sonrisa', 1), ('proced\xada', 1), ('ahora', 1), ('de', 112), ('la', 60), ('idea', 2), ('de', 112), ('su', 19), ('inmolaci\xed', 1), ('un', 4), ('punto', 1), ('d\xedbil', 2), ('ten\xada', 2), ('este', 3), ('Fortunato', 5), ('aunque', 1), ('en', 40), ('otros', 1), ('sentidos', 1), ('era', 9), ('hombre', 1), ('de', 112), ('respetar', 1), ('y', 54), ('aun', 1), ('de', 112), ('temer', 1), ('Enorgullec\xed', 1), ('de', 112), ('ser', 1), ('un', 28), ('connaissanceur', 1), ('en', 40), ('materia', 1), ('de', 112), ('vinos', 1), ('Pocos', 1), ('italianos', 1), ('poseen', 1), ('la', 60), ('capacidad', 1), ('del', 17), ('verdadero', 1), ('virtuoso', 1), ('En', 7), ('su', 19), ('mayor', 1), ('parte', 1), ('el', 38), ('entusiasmo', 1), ('que', 53), ('fingen', 1), ('se', 16), ('adapta', 1), ('al', 10), ('momento', 1), ('y', 54), ('a', 41), ('la', 60), ('oportunidad', 1), ('a', 41), ('fin', 2), ('de', 112), ('enga\xedar', 1), ('a', 41), ('los', 17), ('millonarios', 1), ('ingleses', 1), ('y', 54), ('austriacos', 1), ('En', 7), ('pintura', 1), ('y', 54), ('en', 40), ('alhajas', 1), ('Fortunato', 7), ('era', 9), ('un', 28), ('impostor', 1), ('como', 8), ('todos', 4), ('sus', 10), ('compatriotas', 1), ('pero', 8), ('
```

Fig.2. Numero de palabras en total y ocurrencias de cada una.

Link de GitHub: <https://github.com/AlmDiana/Análisis-de-datos/tree/Ocurrencias-de-palabras>