

# Archivos. Ejercicios

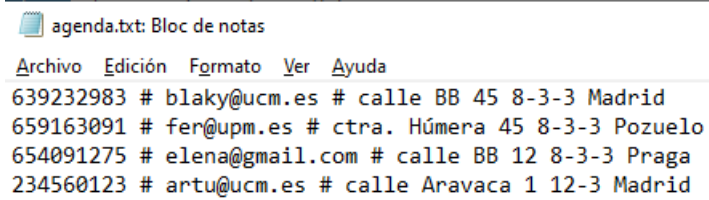
---

## Agenda básica

Tenemos una lista de líneas de texto como la siguiente:

```
"639232983 # blaky@ucm.es # calle BB 45 8-3-3 Madrid"
```

contenida en un archivo de disco:

The image shows a screenshot of a text editor window titled 'agenda.txt: Bloc de notas'. The editor has a menu bar with 'Archivo', 'Edición', 'Formato', 'Ver', and 'Ayuda'. The text content of the file is as follows:

```
639232983 # blaky@ucm.es # calle BB 45 8-3-3 Madrid
659163091 # fer@upm.es # ctra. Húmera 45 8-3-3 Pozuelo
654091275 # elena@gmail.com # calle BB 12 8-3-3 Praga
234560123 # artu@ucm.es # calle Aravaca 1 12-3 Madrid
```

Cada una contiene un número de teléfono, una dirección de email y una dirección física. El separador es la cadena de tres espacios " # " .

Diseña una función que toma como entrada el nombre del archivo de disco en que está almacenada dicha agenda y forma un diccionario, donde la clave es el número de teléfono, y el valor, un nuevo diccionario con dos elementos: el email y la dirección.

Nota: usa el método `split(...)` para descomponer cada línea, y la función `rstrip()` para eliminar las marcas de fin de línea.

In [1]:



```
def cargar_agenda(nombre_archivo):
    agenda = dict({})
    el_archivo = open(nombre_archivo, "r")
    for l in el_archivo:
        datos = l.split(" # ")
        agenda[datos[0]] = {
            "email" : datos[1],
            "direc" : datos[2].rstrip()
        }
    el_archivo.close()
    return agenda

agenda = cargar_agenda("agenda.txt")
print(agenda)

print()

print(agenda["234560123"]["email"])
```

```
{'639232983': {'email': 'blaky@ucm.es', 'direc': 'calle BB 45 8-3-3 Madri
d'}, '659163091': {'email': 'fer@upm.es', 'direc': 'ctra. Húmera 45 8-3-3 Po
zuelo'}, '654091275': {'email': 'elena@gmail.com', 'direc': 'calle BB 12 8-3
-3 Praga'}, '234560123': {'email': 'artu@ucm.es', 'direc': 'calle Aravaca 1
12-3 Madrid'}}
```

artu@ucm.es

## Numerar líneas

Diseña una función que lee las líneas de un archivo como el siguiente:

```
En un lugar de la Mancha
de cuyo nombre no quiero acordarme
vivía un hidalgo
...
```

y las escribe, numerándolas:

```
1. En un lugar de la Mancha
2. de cuyo nombre no quiero acordarme
3. vivía un hidalgo
...
```

In [2]:



```
def numerar_lineas(nombre_archivo):
    el_archivo = open(nombre_archivo, "r")
    num_linea = 0
    for linea in el_archivo:
        num_linea = num_linea + 1
        print(num_linea, linea.rstrip())
        if num_linea == 25: # Para que no sea interminable
            el_archivo.close()
            return

numerar_lineas("quijote - fragmento.txt")
```

1 Primera parte del ingenioso hidalgo don Quijote de la Mancha  
2  
3 Capítulo primero. Que trata de la condición y ejercicio del famoso hidalgo  
4 don Quijote de la Mancha  
5  
6 En un lugar de la Mancha, de cuyo nombre no quiero acordarme, no ha mucho  
7 tiempo que vivía un hidalgo de los de lanza en astillero, adarga antigua,  
8 rocín flaco y galgo corredor. Una olla de algo más vaca que carnero,  
9 salpicón las más noches, duelos y quebrantos los sábados, lantejas los  
10 viernes, algún palomino de añadidura los domingos, consumían las tres  
11 partes de su hacienda. El resto della concluían sayo de velarte, calzas d  
e  
12 velludo para las fiestas, con sus pantuflos de lo mismo, y los días de  
13 entresemana se honraba con su vellorí de lo más fino. Tenía en su casa un  
a  
14 ama que pasaba de los cuarenta, y una sobrina que no llegaba a los veint  
e,  
15 y un mozo de campo y plaza, que así ensillaba el rocín como tomaba la  
16 podadera. Frisaba la edad de nuestro hidalgo con los cincuenta años; era  
de  
17 complexión recia, seco de carnes, enjuto de rostro, gran madrugador y ami  
go  
18 de la caza. Quieren decir que tenía el sobrenombre de Quijada, o Quesada,  
19 que en esto hay alguna diferencia en los autores que deste caso escriben;  
20 aunque, por conjeturas verosímiles, se deja entender que se llamaba  
21 Quejana. Pero esto importa poco a nuestro cuento; basta que en la narraci  
ón  
22 dél no se salga un punto de la verdad.  
23  
24 Es, pues, de saber que este sobredicho hidalgo, los ratos que estaba  
25 ocioso, que eran los más del año, se daba a leer libros de caballerías, c  
on

In [3]:



```
def num_palabras_promedio(nombre_archivo):  
    el_archivo = open(nombre_archivo, "r")  
    num_lineas = 0  
    num_palabras = 0  
    for linea in el_archivo:  
        num_lineas = num_lineas + 1  
        num_palabras = num_palabras + len(linea.split())  
    return num_palabras / num_lineas  
  
print(num_palabras_promedio("quijote - fragmento.txt"))
```

10.185703164845783

In [4]:

```
def limpia_palabra(palabra):
    limpia = palabra
    while limpia != "" and not limpia[0].isalpha():
        limpia = limpia[1:]
    while limpia != "" and not limpia[-1].isalpha():
        limpia = limpia[:-1]
    return limpia.lower()

def diccionario_frecuencias(nombre_archivo):
    el_archivo = open(nombre_archivo, "r")
    dicc = dict({})
    for linea in el_archivo:
        for palabra in linea.split():
            pal = limpia_palabra(palabra)
            if pal in dicc:
                dicc[pal] = dicc[pal] + 1
            else:
                dicc[pal] = 1
    return dicc

def fp(pal_freq):
    dicc = dict({})
    for palabra, freq in pal_freq.items():
        if freq in dicc:
            dicc[freq].append(palabra)
        else:
            dicc[freq] = [palabra]
    return dicc

palabras_frecuencias = diccionario_frecuencias("quijote - fragmento.txt")
for p in ["en", "un", "lugar", "adarga", "flaco", "galgo", "corredor", \
          "mancha", "alonso", "dulcinea", "molino"]:
    print(p, palabras_frecuencias[p])

print()

print("Palabras más frecuentes:")

freq_palabras = fp(palabras_frecuencias)
for f, pals in freq_palabras.items():
    if f > 5000:
        print(f, pals)

print()
print("Palabras que aparecen sólo una vez:")

for f, pals in freq_palabras.items():
    if f == 1:
        print(f, pals)
```

```
en 8109
un 1910
lugar 343
adarga 19
flaco 11
galgo 3
corredor 3
```

mancha 162  
alonso 10  
dulcinea 277  
molino 11

Palabras más frecuentes:

17953 ['de']  
10227 ['la']  
20412 ['que']  
17981 ['y']  
8109 ['en']  
6283 ['no']

