

# Manejador de Archivos como una Secuencia

- Un **manejador de archivos** abierto en modo de lectura puede ser tratado como una **secuencia** de cadenas donde cada línea en el archivo es una cadena en la secuencia
- Podemos usar la sentencia **for** para iterar a través de una **secuencia**
- Recuerda - una **secuencia** es un conjunto ordenado

```
archivo = open('mbox.txt')  
for queso in archivo :  
    print(queso)
```

# Contando Líneas en un Archivo

- Abre un **archivo** en modo de lectura
- Utiliza un bucle **for** para leer cada línea
- **Cuenta** las líneas e imprime el número total de líneas

```
man_a = open('mbox.txt')
contador = 0
for linea in man_a:
    contador = contador + 1
print('Total de Líneas:', contador)
```

```
$ python abrir.py
Total de Líneas: 132045
```

# Leyendo el Archivo \*Entero\*

Podemos leer un archivo entero (saltos de líneas y todo lo demás) dentro de una sola cadena

```
>>> man_a = open('mbox-short.txt')
>>> ent = man_a.read()
>>> print(len(ent))
94626
>>> print(ent[:20])
From stephen.marquar
```

# Búsqueda a Través de un Archivo

Podemos poner una sentencia **if** nuestro bucle **for** para únicamente imprimir líneas que satisfacen cierta característica

```
man_a = open('mbox-short.txt')
for linea in man_a:
    if linea.startswith('From: ') :
        print(linea)
```

# ¡Uy!

¿Qué están haciendo ahí  
todas esas líneas en  
blanco?

```
From: stephen.marquard@uct.ac.za
```

```
From: louis@media.berkeley.edu
```

```
From: zqian@umich.edu
```

```
From: rjlowe@iupui.edu
```

```
...
```

# ¡Uy!

¿Qué están haciendo ahí todas esas líneas en blanco?

- Cada línea del archivo tiene un **salto de línea** al final
- La sentencia **print** agrega un **salto de línea** a cada línea

```
From: stephen.marquard@uct.ac.za\n\nFrom: louis@media.berkeley.edu\n\nFrom: zqian@umich.edu\n\nFrom: rjlowe@iupui.edu\n\n...
```

# Búsqueda a Través de un Archivo (arreglado)

- Podemos remover los espacios en blanco del lado derecho de la cadena utilizando `rstrip()` de la librería de cadenas (string)
- El salto de línea es considerado como un “espacio en blanco” y es **removido**

```
man_a = open('mbox-short.txt')
for linea in man_a:
    linea = linea.rstrip()
    if linea.startswith('From: ') :
        print(linea)
```

```
From: stephen.marquard@uct.ac.za
From: louis@media.berkeley.edu
From: zqian@umich.edu
From: rjlowe@iupui.edu
....
```

# Ignorando con Continue

Podemos ignorar líneas de forma conveniente al utilizar la sentencia **continue** (continuar)


```
man_a = open('mbox-short.txt')
for line in man_a:
    line = line.rstrip()
    if not line.startswith('From: ') :
        continue
    print(line)
```



# Usando **in** para Seleccionar Líneas

Podemos buscar una cadena en cualquier parte **en** una **línea** como nuestro criterio de selección

```
man_a = open('mbox-short.txt')
for line in man_a:
    line = line.rstrip()
    if not '@uct.ac.za' in line:
        continue
    print(line)
```



```
From stephen.marquard@uct.ac.za Sat Jan  5 09:14:16 2008
X-Authentication-Warning: set sender to stephen.marquard@uct.ac.za using -f
From: stephen.marquard@uct.ac.za
Author: stephen.marquard@uct.ac.za
From david.horwitz@uct.ac.za Fri Jan  4 07:02:32 2008
X-Authentication-Warning: set sender to david.horwitz@uct.ac.za using -f...
```

# Solicitar Nombre de Archivo

```
nombrea = input('Ingresa un nombre de archivo:')  
man_a = open(nombrea)  
contador = 0  
for linea in man_a:  
    if linea.startswith('Subject:') :  
        contador = contador + 1  
print('Había', contador, 'líneas de subject en', nombrea)
```

Ingresa un nombre de archivo: mbox.txt  
Había 1797 líneas de subject en mbox.txt

Ingresa un nombre de archivo: mbox-short.txt  
Había 27 líneas de subject en mbox-short.txt

# Nombres de Archivo Incorrectos

```
nombrea = input('Ingresa un nombre de archivo: ')
try:
    man_a = open(nombrea)
except:
    print('El archivo no se puede abrir:', nombrea)
    quit()

contador = 0
for linea in man_a:
    if linea.startswith('Subject:') :
        contador = contador + 1
print('Había', contador, 'líneas de subject en', nombrea)
```

Ingresa un nombre de archivo: mbox.txt  
Había 1797 líneas de subject en mbox.txt

Ingresa un nombre de archivo: na na boo boo  
El archivo no se puede abrir: na na boo boo

# Resumen

- Almacenamiento Secundario
- Abriendo un archivo - manejador de archivo
- Estructura de archivo - salto de línea
- Lectura de un archivo línea por línea con un bucle for
- Búsqueda por líneas
- Leyendo nombres de archivos
- Manejando archivos incorrectos



## Agradecimientos / Contribuciones



Las diapositivas están bajo el Copyright 2010- Charles R. Severance ([www.dr-chuck.com](http://www.dr-chuck.com)) de la Escuela de Informática de la Universidad de Michigan y [open.umich.edu](http://open.umich.edu), y están disponibles públicamente bajo una Licencia Creative Commons Attribution 4.0. Favor de mantener esta última diapositiva en todas las copias del documento para cumplir con los requerimientos de atribución de la licencia. Si haces un cambio, siéntete libre de agregar tu nombre y organización a la lista de contribuidores en esta página conforme sean republicados los materiales.

Desarrollo inicial: Charles Severance, Escuela de Informática de la Universidad de Michigan.

Traducción al Español por Juan Carlos Pérez Castellanos - 2020-04-11