

# Archivos en Internet

Esto no es web-scrapping, pero también podemos descargar archivos que están en Internet desde un programa y procesarlos como los archivos de disco.

Hay varias librerías: `urllib` , `urllib2` , `urllib3` , etc.

In [1]: ▶ *# La librería para acceder a archivos en Internet:*

```
import urllib
```

```
# Elijo el Quijote, cómo no, disponible en la librería Gutenberg:
```

```
url_quijote = "https://www.gutenberg.org/cache/epub/2000/pg2000.txt"
f = urllib.request.urlopen(url_quijote)
```

```
# Ahora f es como un archivo más. Por eso podemos aplicarle la función readlines:
texto_en_bruto = f.readlines()
```

```
for linea in texto_en_bruto[613:620]:
    print(linea)
```

```
b'En un lugar de la Mancha, de cuyo nombre no quiero acordarme, no ha mucho\r\n'
b'tiempo que viv\x3\xada un hidalgo de los de lanza en astillero, adarga antigua,\r\n'
b'roc\x3\xadn flaco y galgo corredor. Una olla de algo m\x3\xa1s vaca que carnero,\r\n'
b'salpic\x3\xb3n las m\x3\xa1s noches, duelos y quebrantos los s\x3\xadados, lantejas los\r\n'
b'viernes, alg\x3\xban palomino de a\x3\xadadidura los domingos, consum\x3\xadan las tres\r\n'
b'partes de su hacienda. El resto della conclu\x3\xadan sayo de velarte, calzas de\r\n'
b'velludo para las fiestas, con sus pantuflos de lo mismo, y los d\x3\xadas de\r\n'
```