

Creado por:

Isabel Maniega

Scrapeo con BeautifulSoup

```
In [3]: import warnings
warnings.filterwarnings('ignore')
```

```
In [1]: # para realizar la petición HTTP y poder leer el documento HTML
# pip install requests
```

```
In [2]: # Para realizar las extracciones de las diferentes partes que nos interesan
# pip install beautifulsoup4
```

```
In [5]: import requests
from bs4 import BeautifulSoup
```

```
In [8]: page = requests.get('https://parascrapear.com/')
print(page)
print('\n')
soup = BeautifulSoup(page.text, 'html.parser')
print(soup)
```

<Response [200]>

```
<html>
<head>
<title>ParaScrapear.com</title>
<meta content="width=device-width, initial-scale=1.0" name="viewport"/>
<link href="/style.css?v=8" rel="stylesheet"/>
</head>
<body>
<h1><a href="/">ParaScrapear.com</a><code>&lt;HTML&gt;</code></h1>
<div id="toptext">
<p style="text-align: center; margin: 40px auto;"><a href="/login">Iniciar sesión</a> | <a href="/javascript">JavaScript</a></p>
</div>
<blockquote>
<p><q>Prefiero equivocarme creyendo en un Dios que no existe, que equivocarme no creyendo en un Dios que exist
e. Porque si después no hay nada, evidentemente nunca lo sabré, cuando me hunda en la nada eterna; pero si hay
algo, si hay Alguien, tendré que dar cuenta de mi actitud de rechazo.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-blaise-pascal">Blaise Pascal</a></footer>
</blockquote>
<blockquote>
<p><q>El hombre ha nacido libre y por doquiera se encuentra sujeto con cadenas.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-jean-jacques-rousseau">Jean Jacques Rousseau</a></footer>
</blockquote>
<blockquote>
<p><q>Nuestra naturaleza está en movimiento. El reposo absoluto es la muerte.</q></p>
<p><a class="cat" href="frases-de-naturaleza">Naturaleza</a></p>
<footer><a class="author" href="frases-de-blaise-pascal">Blaise Pascal</a></footer>
</blockquote>
<blockquote>
<p><q>Quisiera sufrir todas las humillaciones, todas las torturas, el ostracismo absoluto y hasta la muerte, pa
ra impedir la violencia.</q></p>
<p><a class="cat" href="frases-de-defectos">Defectos</a></p>
<footer><a class="author" href="frases-de-mahatma-gandhi">Mahatma Gandhi</a></footer>
</blockquote>
<p><q>La única simplicidad que vale la pena de conservar es la del corazón, la simplicidad que acepta y goza.</
q></p>
<p><a class="cat" href="frases-de-miscelanea">Miscelánea</a></p>
<footer><a class="author" href="frases-de-gilbert-keith-chesterton">Gilbert Keith Chesterton</a></footer>
</blockquote>
<p><q>Las ideas audaces son como piezas de ajedrez. Pueden ser vencidas, pero también pueden iniciar una partid
a victoriosa.</q></p>
<p><a class="cat" href="frases-de-pensamiento-y-razon">Pensamiento y razón</a></p>
<footer><a class="author" href="frases-de-goethe">Goethe</a></footer>
</blockquote>
<p><q>No me dejes caer en el vulgar error de soñar que soy perseguido cada vez que alguien me contradice.</q></
p>
<p><a class="cat" href="frases-de-miscelanea">Miscelánea</a></p>
<footer><a class="author" href="frases-de-emerson">Emerson</a></footer>
</blockquote>
<p><q>Un político divide a las personas en dos grupos: en primer lugar, instrumentos; en segundo, enemigos.</q>
</p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-friedrich-nietzsche">Friedrich Nietzsche</a></footer>
</blockquote>
<p><q>El más importante y principal negocio público es la buena educación de la juventud.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-platon">Platón</a></footer>
</blockquote>
<blockquote>
<p><q>La literatura no es otra cosa que un sueño dirigido.</q></p>
<p><a class="cat" href="frases-de-arte">Arte</a></p>
<footer><a class="author" href="frases-de-jorge-luis-borges">Jorge Luis Borges</a></footer>
</blockquote>
<p><q>¿Cuál es la tarea más difícil del mundo? Pensar.</q></p>
<p><a class="cat" href="frases-de-pensamiento-y-razon">Pensamiento y razón</a></p>
<footer><a class="author" href="frases-de-emerson">Emerson</a></footer>
</blockquote>
<p><q>Se piensa que lo justo es lo igual, y así es; pero no para todos, sino para los iguales. Se piensa por el
contrario que lo justo es lo desigual, y así es, pero no para todos, sino para los desiguales.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-aristoteles">Aristóteles</a></footer>
</blockquote>
<p><q>Saber olvidar, más es dicha que arte.</q></p>
<p><a class="cat" href="frases-de-miscelanea">Miscelánea</a></p>
<footer><a class="author" href="frases-de-baltasar-gracian">Baltasar Gracián</a></footer>
</blockquote>
<p><q>Bueno es tener la alegría en casa y no haber menester de buscarla fuera.</q></p>
<p><a class="cat" href="frases-de-sentimientos">Sentimientos</a></p>
<footer><a class="author" href="frases-de-goethe">Goethe</a></footer>
</blockquote>
<blockquote>
<p><q>No hace falta un gobierno perfecto; se necesita uno que sea práctico.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-aristoteles">Aristóteles</a></footer>
</blockquote>
<p><q>Los hombres son criaturas muy raras: la mitad censura lo que practica; la otra mitad practica lo que cens
ura; el resto siempre dice y hace lo que debe.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-benjamin-franklin">Benjamin Franklin</a></footer>
</blockquote>
<p><q>Conviene siempre esforzarse más en ser interesante que exacto; porque el espectador lo perdona todo menos
el sopor.</q></p>
<p><a class="cat" href="frases-de-sentimientos">Sentimientos</a></p>
<footer><a class="author" href="frases-de-voltaire">Voltaire</a></footer>
</blockquote>
<p><q>El peor pecado hacia nuestros semejantes no es odiarlos, sino tratarlos con indiferencia: esa es la esenc
ia de la inhumanidad.</q></p>
<p><a class="cat" href="frases-de-sentimientos">Sentimientos</a></p>
<footer><a class="author" href="frases-de-george-bernard-shaw">George Bernard Shaw</a></footer>
</blockquote>
<p><q>La vida es una serie de colisiones con el futuro; no es una suma de lo que hemos sido, sino de lo que anh
elamos ser.</q></p>
<p><a class="cat" href="frases-de-naturaleza">Naturaleza</a></p>
<footer><a class="author" href="frases-de-jose-ortega-y-gasset">José Ortega y Gasset</a></footer>
</blockquote>
<p><q>Dios es la evidencia invisible.</q></p>
<p><a class="cat" href="frases-de-sociedad">Sociedad</a></p>
<footer><a class="author" href="frases-de-victor-hugo">Victor Hugo</a></footer>
</blockquote>
<div style="position: relative; margin: 40px 0px"><a class="next" href="/index-2" style="position: absolute; ri
ght: 0px;">Siguiente </a></div>
<div id="modals"></div>
<div class="copyright">Hecho con ♥ por <a href="https://scraping.link/">Scraping.link</a></div>
<script src="/app.js?v=2"></script>
</body>
</html>
```

```
In [11]: blockquote_items = soup.find_all('blockquote')
# print(blockquote_items)

for blockquote in blockquote_items:
    autor = blockquote.find(class_='author').text
    categoria = blockquote.find(class_='cat').text
    frase = blockquote.find('q').text
    print([autor, categoria, frase])
    print('\n')
```

['Blaise Pascal', 'Sociedad', 'Prefiero equivocarme creyendo en un Dios que no existe, que equivocarme no creye en un Dios que existe. Porque si después no hay nada, evidentemente nunca lo sabré, cuando me hunda en la nada eterna; pero si hay algo, si hay Alguien, tendré que dar cuenta de mi actitud de rechazo.']

['Jean Jacques Rousseau', 'Sociedad', 'El hombre ha nacido libre y por doquiera se encuentra sujeto con cadena s.']

['Blaise Pascal', 'Naturaleza', 'Nuestra naturaleza está en movimiento. El reposo absoluto es la muerte.']

['Mahatma Gandhi', 'Defectos', 'Quisiera sufrir todas las humillaciones, todas las torturas, el ostracismo abso luto y hasta la muerte, para impedir la violencia.']

['Gilbert Keith Chesterton', 'Miscelánea', 'La única simplicidad que vale la pena de conservar es la del corazón n, la simplicidad que acepta y goza.']

['Goethe', 'Pensamiento y razón', 'Las ideas audaces son como piezas de ajedrez. Pueden ser vencidas, pero tamb ién pueden iniciar una partida victoriosa.']

['Emerson', 'Miscelánea', 'No me dejes caer en el vulgar error de soñar que soy perseguido cada vez que alguien me contradice.']

['Friedrich Nietzsche', 'Sociedad', 'Un político divide a las personas en dos grupos: en primer lugar, instrume ntos; en segundo, enemigos.']

['Platón', 'Sociedad', 'El más importante y principal negocio público es la buena educación de la juventud.']

['Jorge Luis Borges', 'Arte', 'La literatura no es otra cosa que un sueño dirigido.']

['Emerson', 'Pensamiento y razón', '¿Cuál es la tarea más difícil del mundo? Pensar.']

['Aristóteles', 'Sociedad', 'Se piensa que lo justo es lo igual, y así es; pero no para todos, sino para los ig uales. Se piensa por el contrario que lo justo es lo desigual, y así es, pero no para todos, sino para los desi guales.']

['Baltasar Gracián', 'Miscelánea', 'Saber olvidar, más es dicha que arte.']

['Goethe', 'Sentimientos', 'Bueno es tener la alegría en casa y no haber menester de buscarla fuera.']

['Aristóteles', 'Sociedad', 'No hace falta un gobierno perfecto; se necesita uno que sea práctico.']

['Benjamin Franklin', 'Sociedad', 'Los hombres son criaturas muy raras: la mitad censura lo que practica; la ot ra mitad practica lo que censura; el resto siempre dice y hace lo que debe.']

['Voltaire', 'Sentimientos', 'Conviene siempre esforzarse más en ser interesante que exacto; porque el espectad or lo perdona todo menos el sopor.']

['George Bernard Shaw', 'Sentimientos', 'El peor pecado hacia nuestros semejantes, no es odiarlos, sino tratarlo s con indiferencia: esa es la esencia de la inhumanidad.']

['José Ortega y Gasset', 'Naturaleza', 'La vida es una serie de colisiones con el futuro; no es una suma de lo que hemos sido, sino de lo que anhelamos ser.']

['Victor Hugo', 'Sociedad', 'Dios es la evidencia invisible.']

Trabajar la información en formato Dataframe

```
In [12]: import pandas as pd
```

```
In [13]: df = pd.DataFrame([], columns=["autor", "categoria", "frase"])
df.head()
```

```
Out[13]:   autor  categoria  frase
```

```
In [14]: blockquote_items = soup.find_all('blockquote')
# print(blockquote_items)

for blockquote in blockquote_items:
    nueva_fila = {"autor": blockquote.find(class_='author').text,
                  "categoria": blockquote.find(class_='cat').text,
                  "frase": blockquote.find('q').text}
    df = df.append(nueva_fila, ignore_index = True)
df
```

```
Out[14]:   autor  categoria  frase
0      Blaise Pascal      Sociedad      Prefiero equivocarme creyendo en un Dios que n...
1  Jean Jacques Rousseau      Sociedad      El hombre ha nacido libre y por doquiera se en...
2      Blaise Pascal      Naturaleza      Nuestra naturaleza está en movimiento. El repo...
3  Mahatma Gandhi      Defectos      Quisiera sufrir todas las humillaciones, todas...
4  Gilbert Keith Chesterton      Miscelánea      La única simplicidad que vale la pena de conse...
5      Goethe      Pensamiento y razón      Las ideas audaces son como piezas de ajedrez. ...
6      Emerson      Miscelánea      No me dejes caer en el vulgar error de soñar q...
7  Friedrich Nietzsche      Sociedad      Un político divide a las personas en dos grupo...
8      Platón      Sociedad      El más importante y principal negocio público ...
9      Jorge Luis Borges      Arte      La literatura no es otra cosa que un sueño dir...
10      Emerson      Pensamiento y razón      ¿Cuál es la tarea más difícil del mundo? Pensar.
11  Aristóteles      Sociedad      Se piensa que lo justo es lo igual, y así es; ...
12  Baltasar Gracián      Miscelánea      Saber olvidar, más es dicha que arte.
13      Goethe      Sentimientos      Bueno es tener la alegría en casa y no haber m...
14  Aristóteles      Sociedad      No hace falta un gobierno perfecto; se necesi...
15  Benjamin Franklin      Sociedad      Los hombres son criaturas muy raras: la mitad ...
16      Voltaire      Sentimientos      Conviene siempre esforzarse más en ser interes...
17  George Bernard Shaw      Sentimientos      El peor pecado hacia nuestros semejantes no es...
18  José Ortega y Gasset      Naturaleza      La vida es una serie de colisiones con el futu...
19      Victor Hugo      Sociedad      Dios es la evidencia invisible.
```

```
In [15]: len(df)
```

```
Out[15]: 20
```

Creado por:

Isabel Maniega