

# m05\_05\_beautiful\_soup\_teoría

April 5, 2021

## 1 Curso de Python do DS ao DEV

Comunidade DS - Meigarom Lopes

## 2 Modulo 05 - Extração de Dados em HTML

## 3 A biblioteca BeautifulSoup - Teoria

```
[103]: from bs4 import BeautifulSoup
```

```
[102]: html_doc = """<html><head><title>The Dormouse's story</title></head>
<body>
<p class="title"><b>The Dormouse's story</b></p>

<p class="story">Once upon a time there were three little sisters; and their_
↳names were
<a href="http://example.com/elsie" class="sister" id="link1">Elsie</a>,
<a href="http://example.com/lacie" class="sister" id="link2">Lacie</a> and
<a href="http://example.com/tillie" class="sister" id="link3">Tillie</a>;
and they lived at the bottom of a well.</p>

<p class="story">...</p>
"""
```

```
[104]: soup = BeautifulSoup( html_doc, 'html.parser' )
```

```
[105]: soup
```

```
[105]: <html><head><title>The Dormouse's story</title></head>
<body>
<p class="title"><b>The Dormouse's story</b></p>
<p class="story">Once upon a time there were three little sisters; and their
names were
<a class="sister" href="http://example.com/elsie" id="link1">Elsie</a>,
<a class="sister" href="http://example.com/lacie" id="link2">Lacie</a> and
<a class="sister" href="http://example.com/tillie" id="link3">Tillie</a>;
and they lived at the bottom of a well.</p>
```

```
<p class="story">...</p>
</body></html>
```

```
[114]: print( soup.body)
```

```
<body>
<p class="title"><b>The Dormouse's story</b></p>
<p class="story">Once upon a time there were three little sisters; and their
names were
<a class="sister" href="http://example.com/elsie" id="link1">Elsie</a>,
<a class="sister" href="http://example.com/lacie" id="link2">Lacie</a> and
<a class="sister" href="http://example.com/tillie" id="link3">Tillie</a>;
and they lived at the bottom of a well.</p>
<p class="story">...</p>
</body>
```

```
[130]: soup.find_all( 'a', id='link1' )[0]
```

```
[130]: <a class="sister" href="http://example.com/elsie" id="link1">Elsie</a>
```