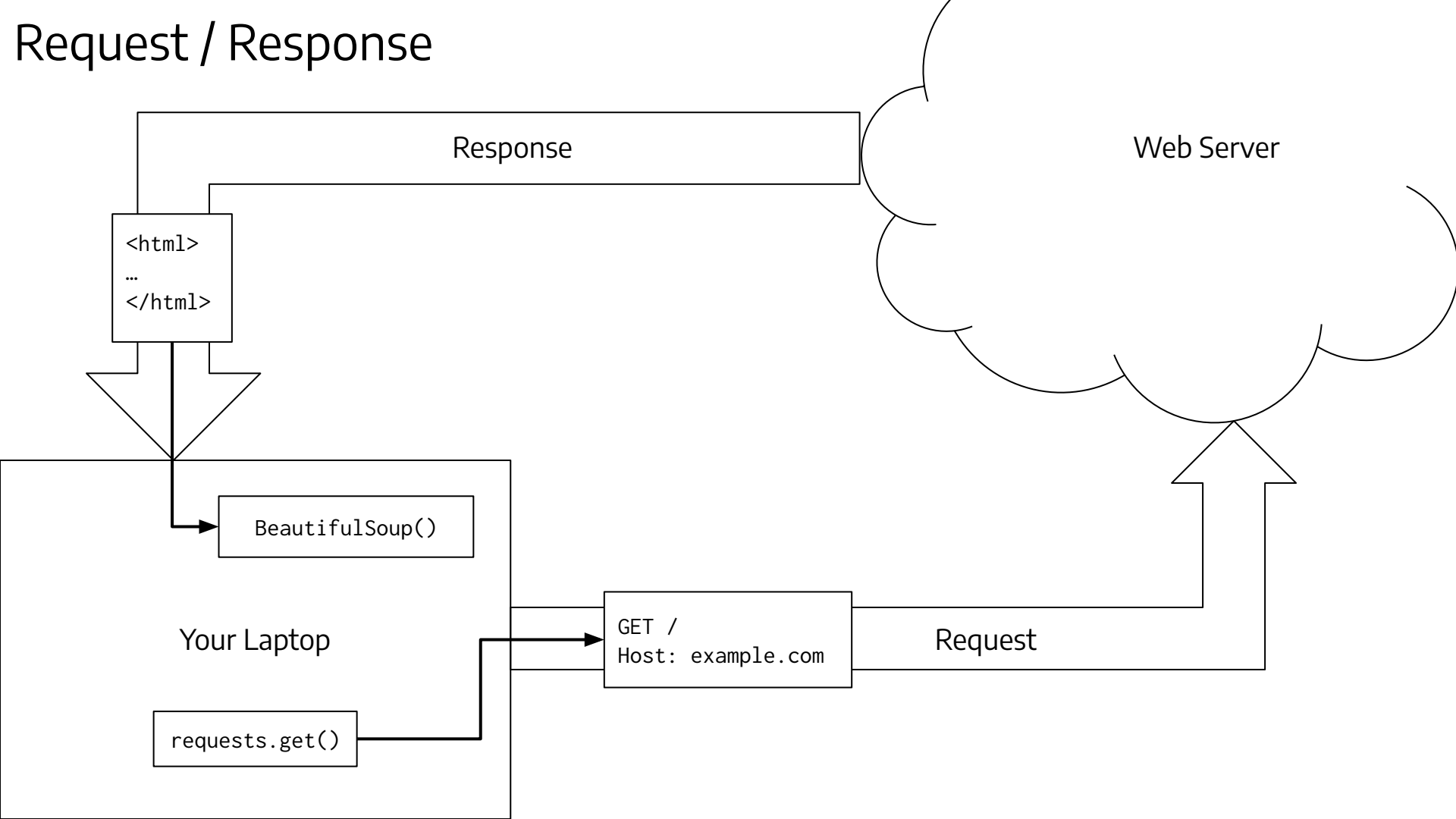
Web Scraping 101

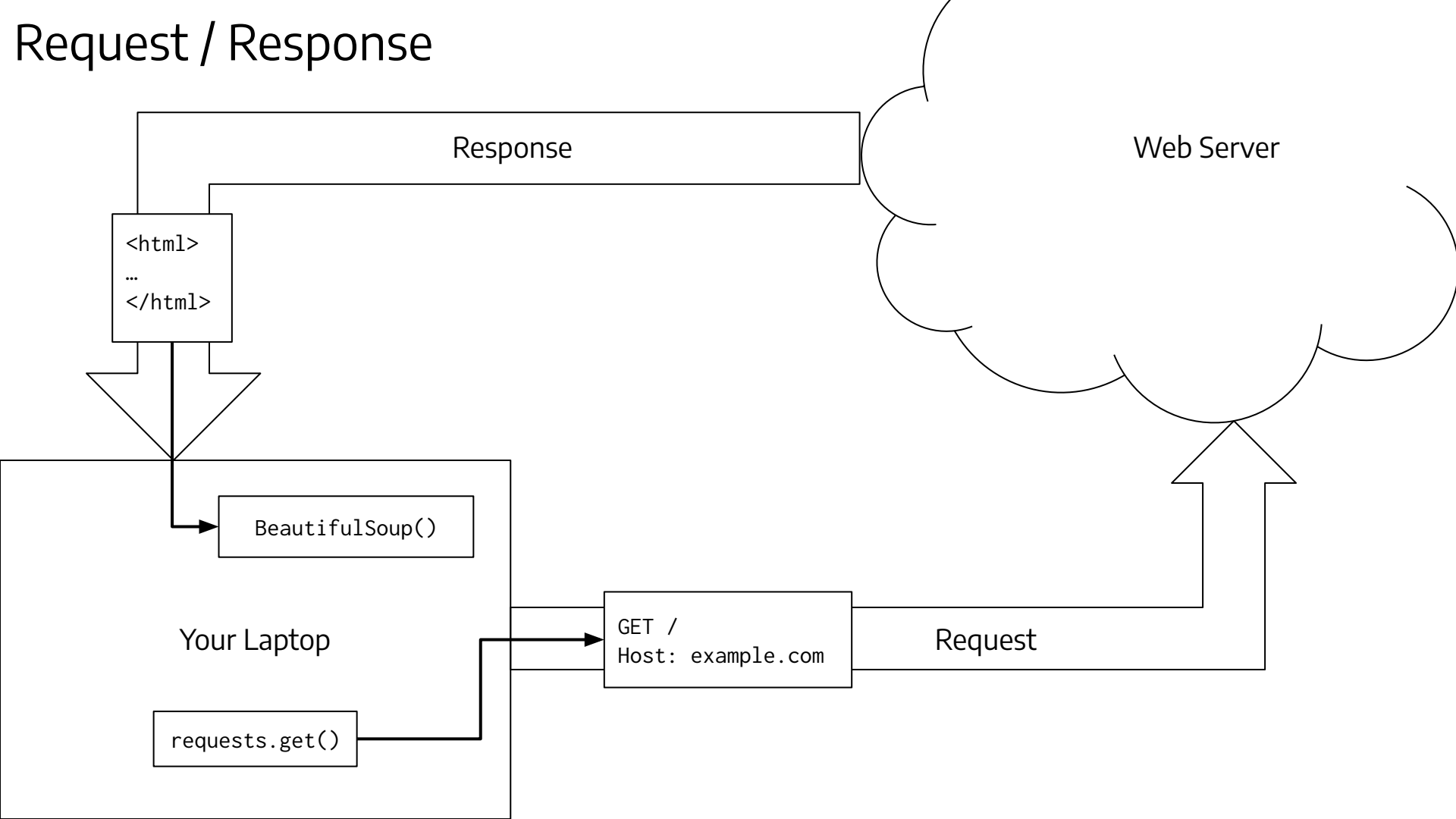
Zach Gulde
2021-10-25

Request / Response

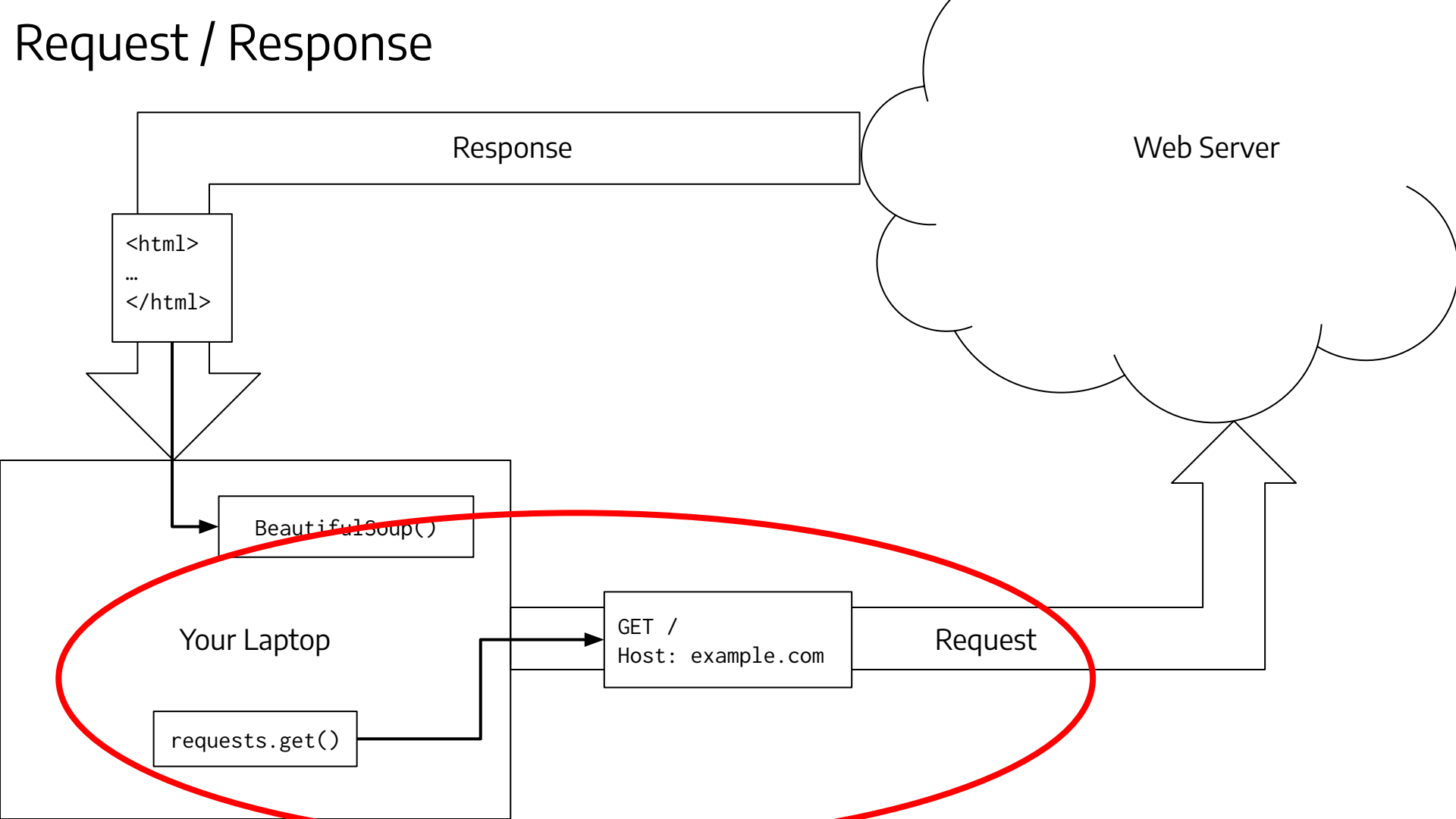


Making a Request

Request / Response



Request / Response



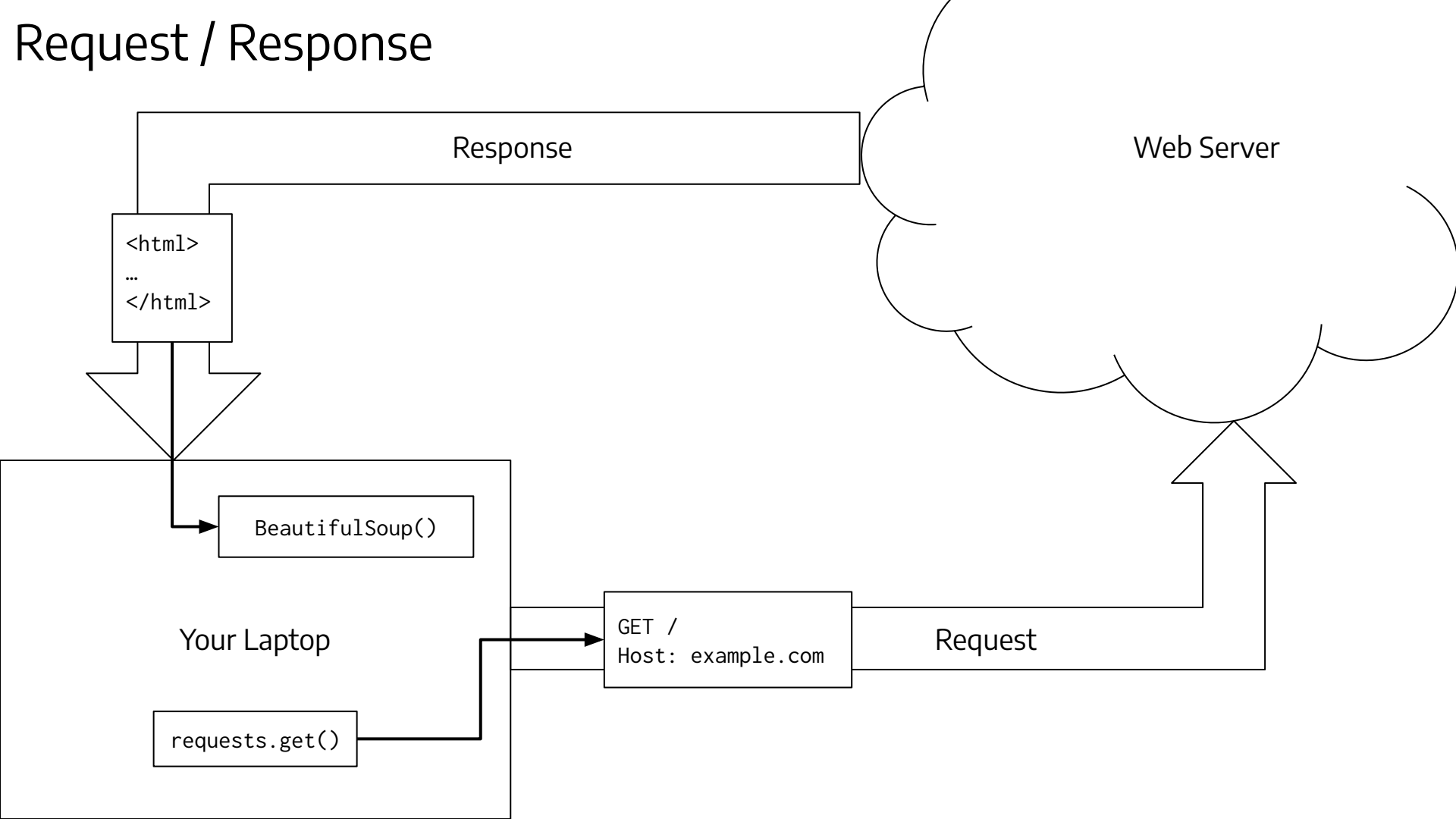
Making a Request

```
import requests

response = requests.get('http://example.com')

response.text
```

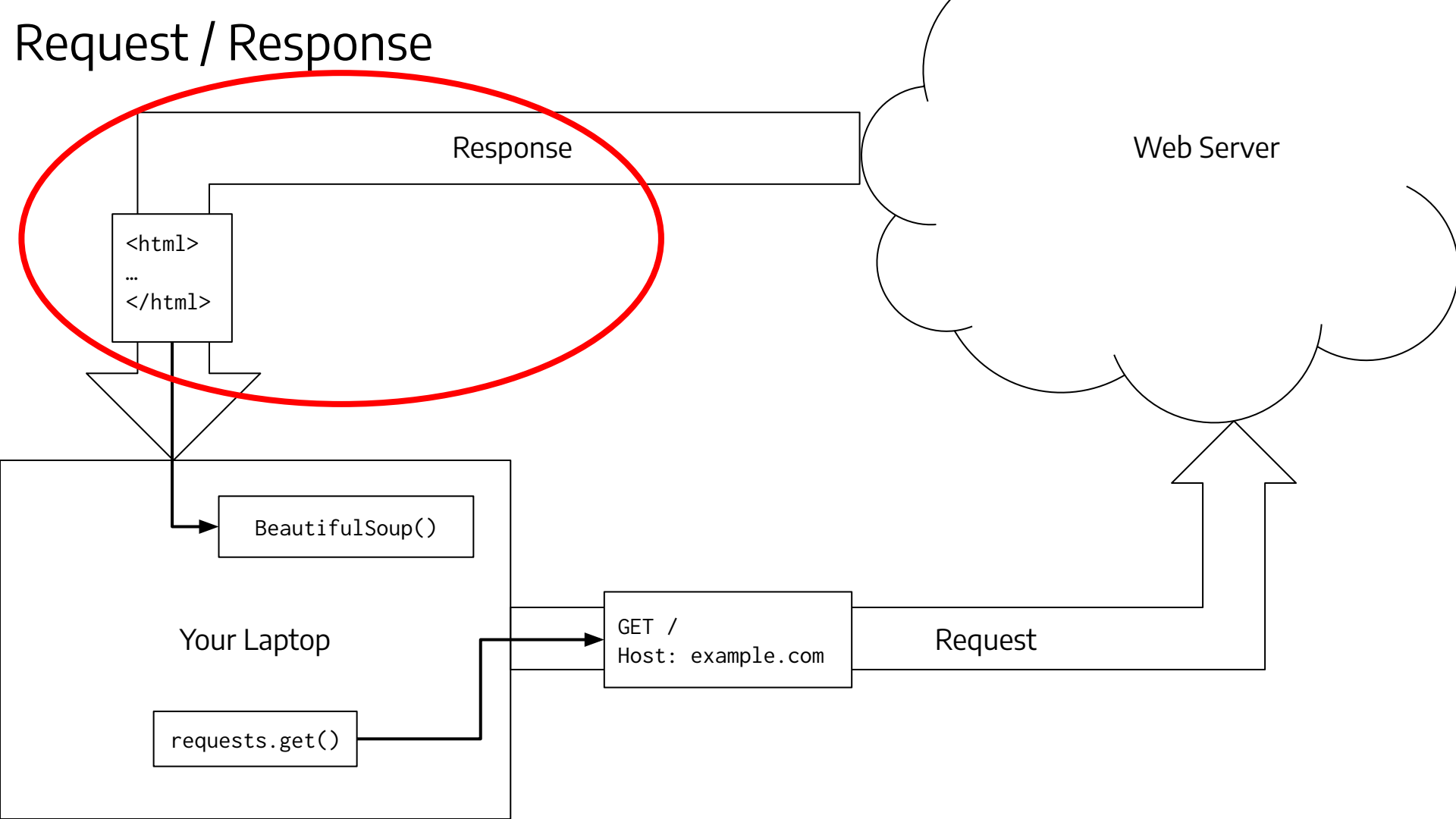
- requests is the library we use to make http requests
- 'http://example.com' is the string argument passed to the .get method; where we are making the request to
- response is an object that holds the response from the server
- response.text gives us access to the raw text of the response body

HTML Documents

Request / Response



Request / Response



HTML Document Anatomy

```
<!DOCTYPE html>
<html lang="en">
<head>
      <meta charset="UTF-8">
      <meta http-equiv="X-UA-Compatible" content="IE=edge">
      <meta name="viewport" content="width=device-width, initial-scale=1.0">
      <title>Web Scraping Demo Pages</title>
      <link href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css" rel="stylesheet" />
      <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.4.1/font/bootstrap-icons.css" />
</head>
<body class="grid justify-center items-center h-screen">
      <div class="text-center space-y-16">
      <h1 class="text-4xl">Web Scraping Demo Pages</h1>
      <div class="grid grid-cols-2 text-2xl">
            <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
            <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
      </div>
      </div>
</body>
</html>
```

HTML Document Anatomy

```
<!DOCTYPE html>
<html lang="en">
<head>
      <meta charset="UTF-8">
      <meta http-equiv="X-UA-Compatible" content="IE=edge">
      <meta name="viewport" content="width=device-width, initial-scale=1.0">
      <title>Web Scraping Demo Pages</title>
      <link href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css" rel="stylesheet" />
      <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.4.1/font/bootstrap-icons.css" />
</head>
<body class="grid justify-center items-center h-screen">
      <div class="text-center space-y-16">
      <h1 class="text-4xl">Web Scraping Demo Pages</h1>
      <div class="grid grid-cols-2 text-2xl">
            <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
            <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
      </div>
      </div>
</body>
</html>
```

Head: Meta Page Information

HTML Document Anatomy

```
<!DOCTYPE html>
<html lang="en">
<head>
      <meta charset="UTF-8">
      <meta http-equiv="X-UA-Compatible" content="IE=edge">
      <meta name="viewport" content="width=device-width, initial-scale=1.0">
      <title>Web Scraping Demo Pages</title>
      <link href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css" rel="stylesheet" />
      <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.4.1/font/bootstrap-icons.css" />
</head>
<body class="grid justify-center items-center h-screen">
      <div class="text-center space-y-16">
      <h1 class="text-4xl">Web Scraping Demo Pages</h1>
      <div class="grid grid-cols-2 text-2xl">
            <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
            <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
      </div>
      </div>
</body>
</html>
```

Head: Meta Page Information

HTML Document Anatomy

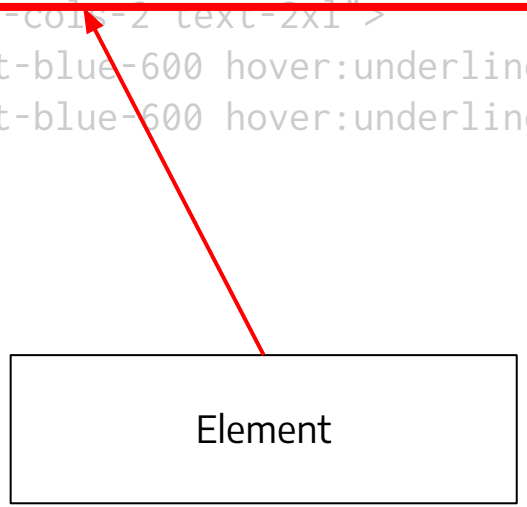
```
<!DOCTYPE html>
<html lang="en">
<head>
      <meta charset="UTF-8">
      <meta http-equiv="X-UA-Compatible" content="IE=edge">
      <meta name="viewport" content="width=device-width, initial-scale=1.0">
      <title>Web Scraping Demo Pages</title>
      <link href="https://unpkg.com/tailwindcss@^2/dist/tailwind.min.css" rel="stylesheet" />
      <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.4.1/font/bootstrap-icons.css" />
</head>
<body class="grid justify-center items-center h-screen">
      <div class="text-center space-y-16">
      <h1 class="text-4xl">Web Scraping Demo Pages</h1>
      <div class="grid grid-cols-2 text-2xl">
            <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
            <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
      </div>
      </div>
</body>
</html>
```

HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

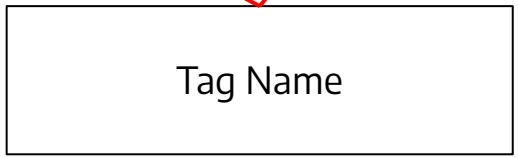
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



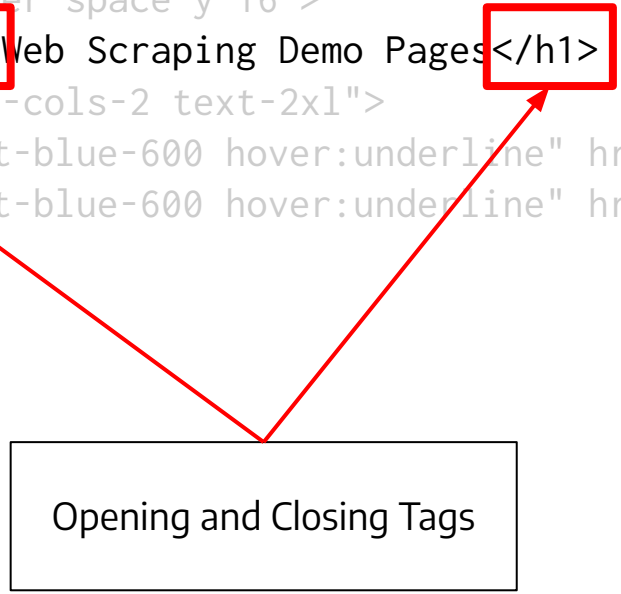
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



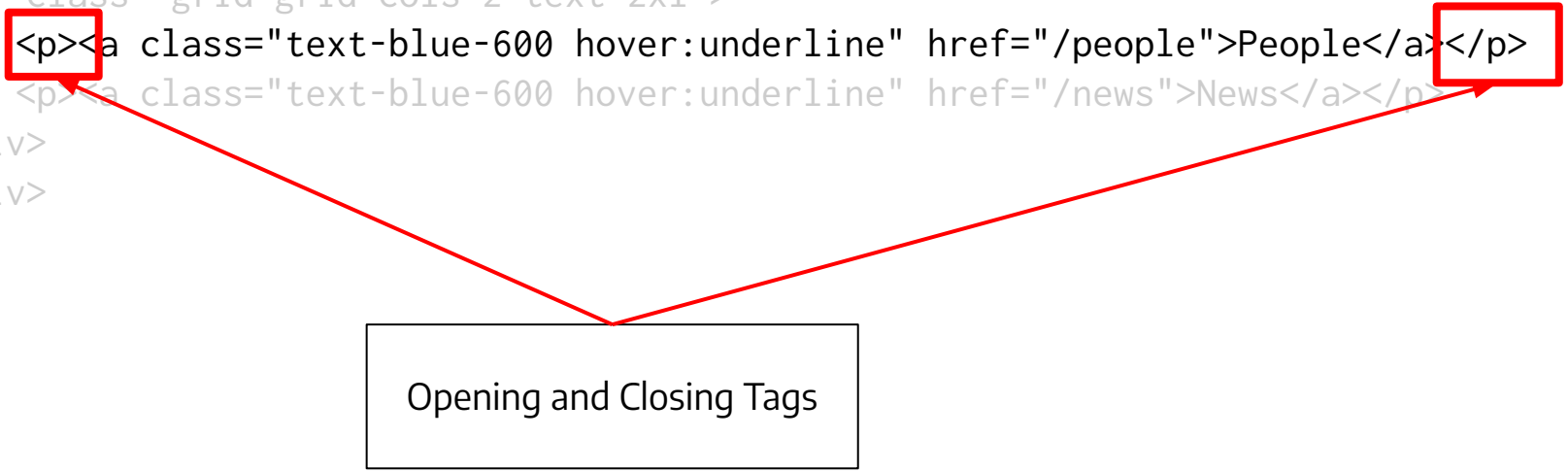
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



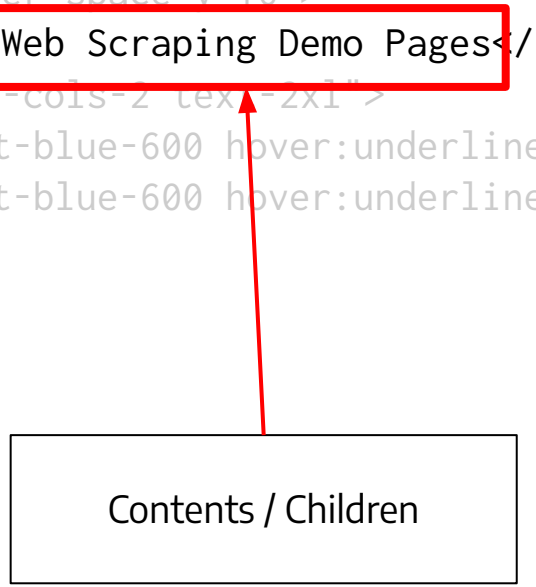
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



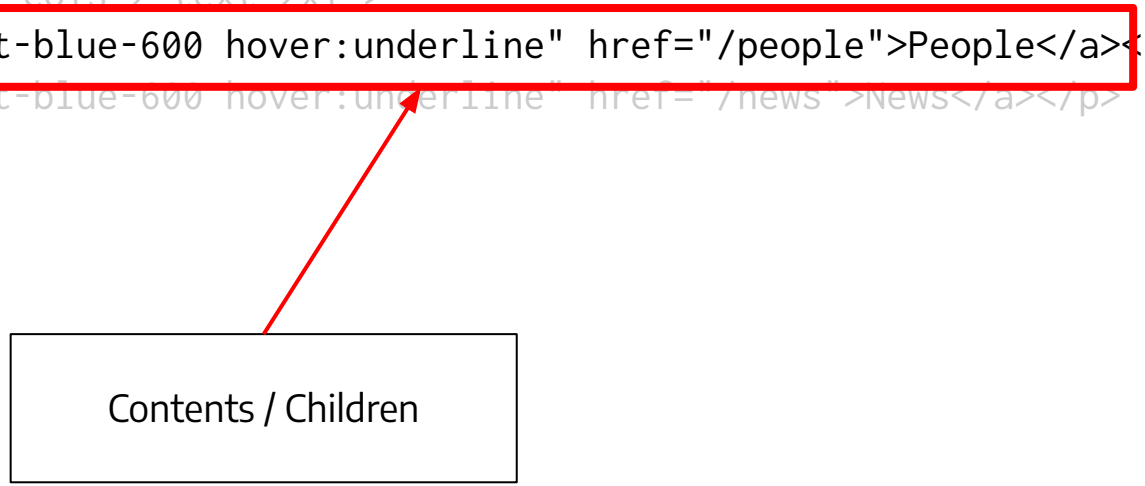
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



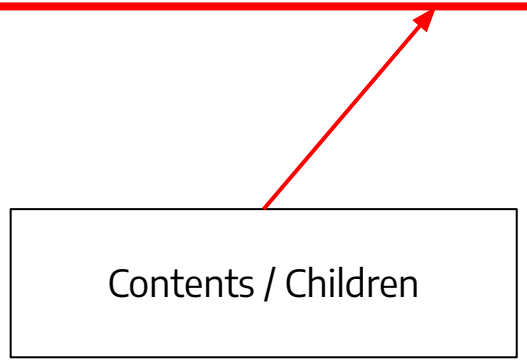
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

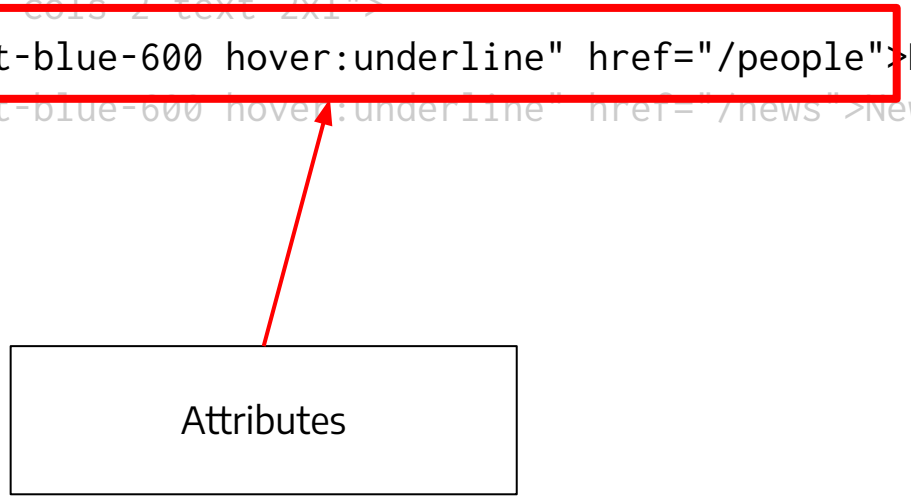


HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

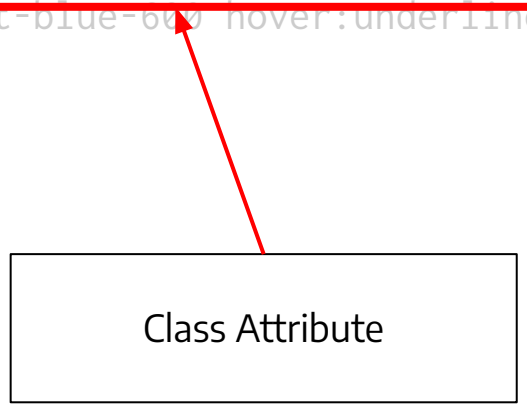
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



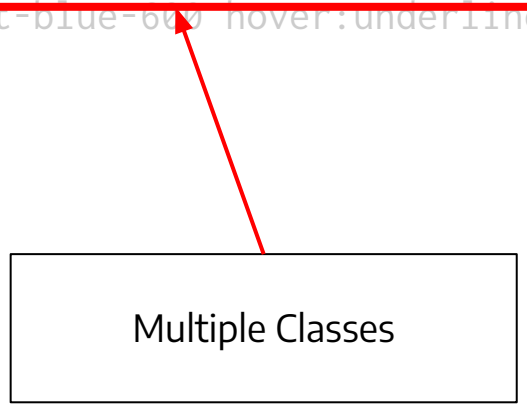
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



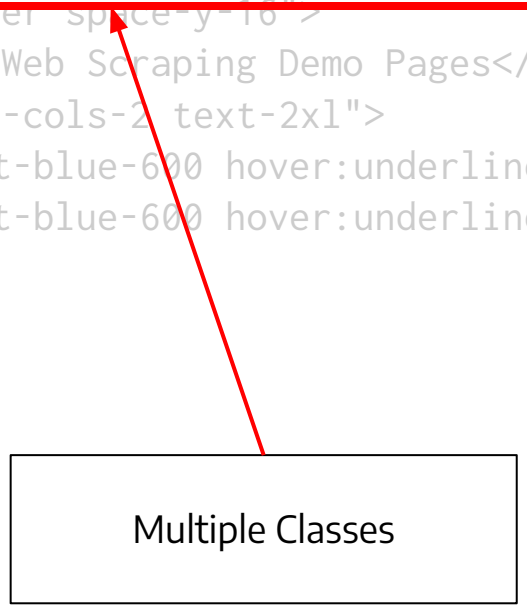
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



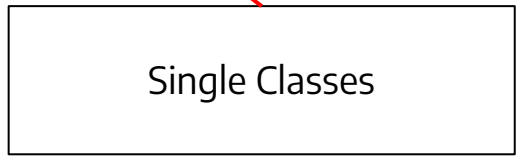
HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



HTML Document Anatomy

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```



HTML Document Anatomy

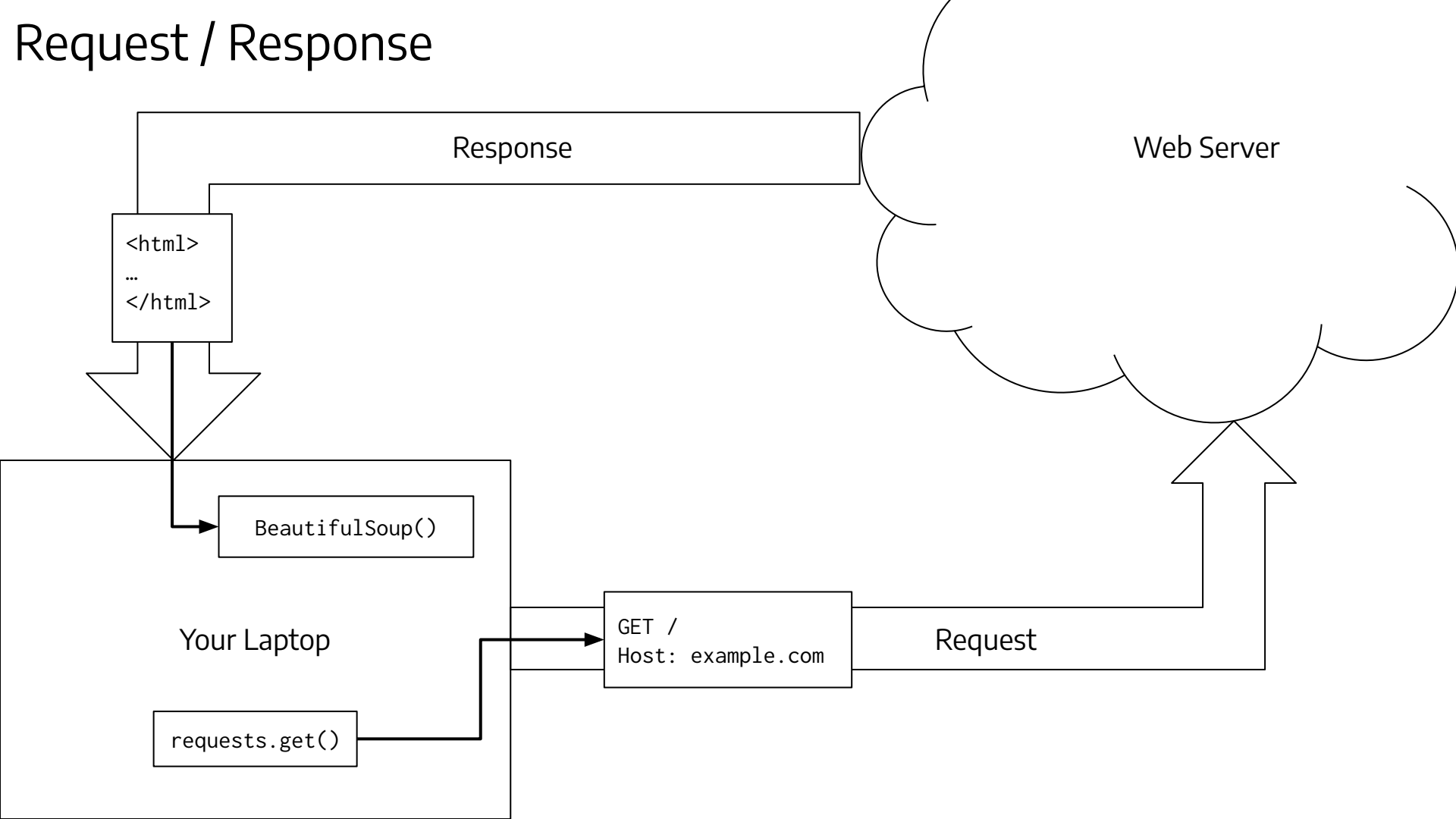
```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Mini Exercise: Describing HTML Documents

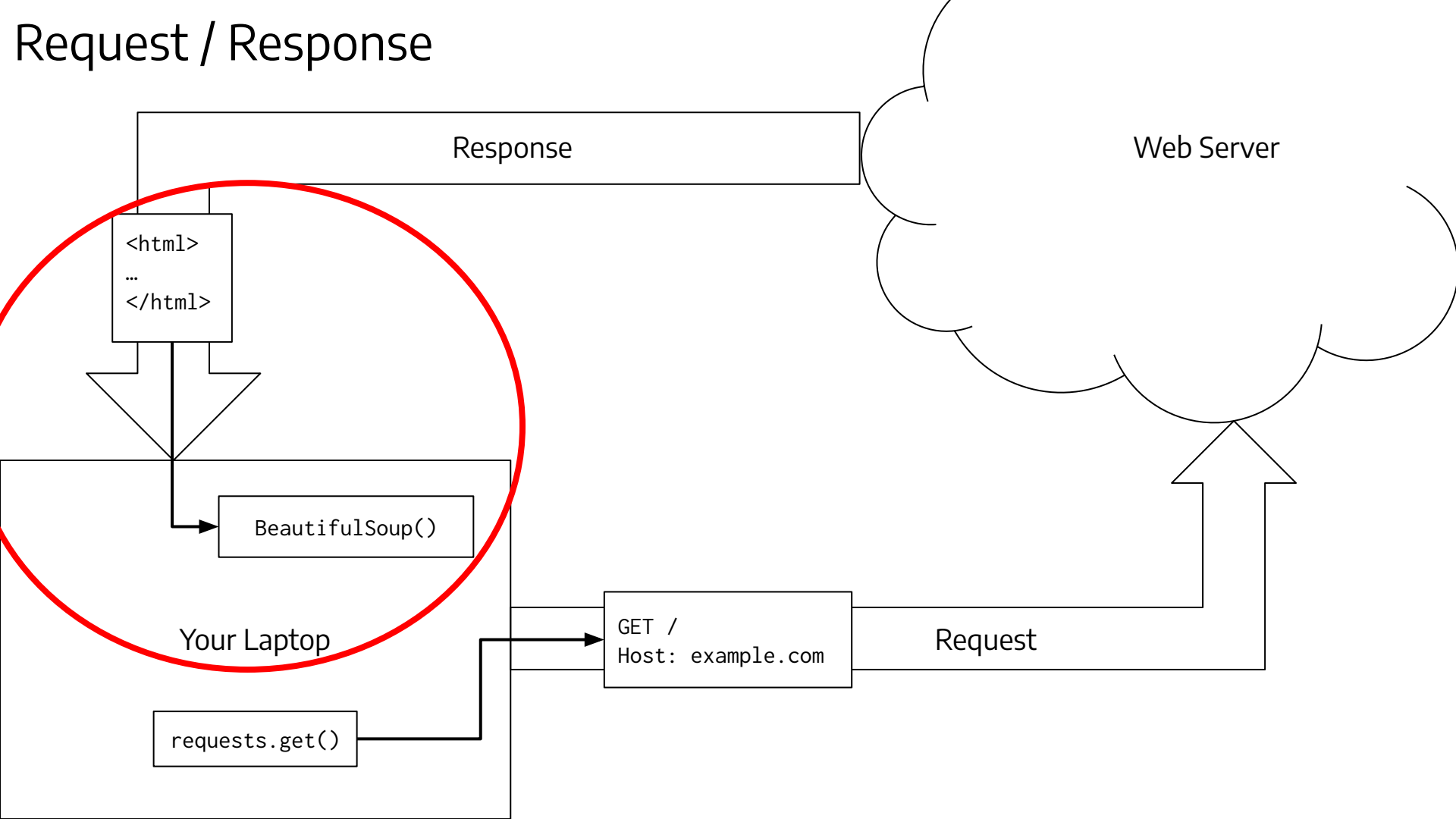
1. View the HTML source for http://example.com.
2. Identify the document head and document body.
3. How many elements are in the document body?
4. What element attributes do you see?

Parsing HTML: Beautiful Soup

Request / Response



Request / Response



Creating a soup object

```
from bs4 import BeautifulSoup

html = '''
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
'''
soup = BeautifulSoup(html)
```

Creating a soup object

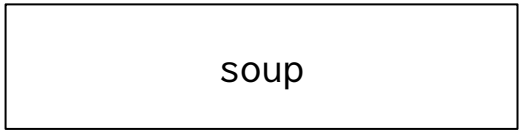
Return a single `Tag` object

- `soup.tag:` find the first element based on tag name
- `soup.find:` find the first element based on tag name or other attributes

Return a list of `Tag` objects

- `soup.findall:` find all elements based on tag name or other attributes
- `soup.selector:` find all elements based on a CSS selector

Using the soup




```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
soup.h1
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
soup.find('h1')
```

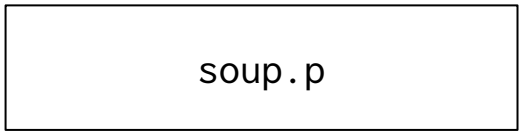
```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
el = soup.find('h1')
el.text  ## "Web Scraping Demo Pages"
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup




```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
soup.find('p')
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
soup.find_all('p')
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
[p.text for p in soup.find_all('p')]  ## ["People", "News"]
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid justify-center items-center h-screen">
    <div class="text-center space-y-16">
    <h1 class="text-4xl">Web Scraping Demo Pages</h1>
    <div class="grid grid-cols-2 text-2xl">
        <p><a class="text-blue-600 hover:underline" href="/people">People</a></p>
        <p><a class="text-blue-600 hover:underline" href="/news">News</a></p>
    </div>
    </div>
</body>
```

Using the soup

```
def extract_links_and_text(el):
    return dict(link=el.attrs['href'], text=el.text)

data = [extract_links_and_text(el) for el in soup.find_all('a')]
pd.DataFrame(data)
```

```
<body class="grid just
    <div class="text-c
    <h1 class="text-4x
    <div class="grid g
        <p><a class="                                    '/people">People</a></p>
        <p><a class="                                    '/news">News</a></p>
    </div>
    </div>
</body>
```

	link	text
0	/people	People
1	/news	News

Demo: scraping news articles

```
…

<body class="mx-auto max-w-screen-lg pb-32">

<h1 class="my-5 text-4xl text-center">News!</h1>
<div class="my-5 text-red-800 px-5 py-3 bg-red-100 font-bold">
    <p>
        <i class="bi bi-exclamation-circle text-xl"></i>
        All data on this page is strictly for demonstration purposes and fake.
    </p>
</div>
<div class="grid gap-y-12">

    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3 bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">peace gas top</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1987-06-04 </p>
                <p class="text-right">By William Stewart </p>
            </div>
            <p>Base impact heart few kitchen sound husband. Fall glass heart section star if.
Trip successful outside audience by. House subject past should.</p>
        </div>
    </div>

    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3 bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">my receive speech</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1990-03-29 </p>
                <p class="text-right">By Steven Moreno </p>
            </div>
            <p>Make education common. Without give talk time always. White ever you must soldier eye.
Particularly course ago bad near guess. Machine blood write mind young behavior. To really hear finally go.</p>
        </div>
    </div>

    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3 bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">good fact then</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1986-11-05 </p>
                <p class="text-right">By Christopher Mason </p>
            </div>
            <p>Rate discuss house back table. Base myself world young leader mother.
Son employee compare pick reality specific. Could course wrong race. Allow standard technology real maintain add.</p>
        </div>
    </div>

...
```

Notice the repeating structure

```
…

<body class="mx-auto max-w-screen-lg pb-32">

<h1 class="my-5 text-4xl text-center">News!</h1>
<div class="my-5 text-red-800 px-5 py-3 bg-red-100 font-bold">
    <p>
        <i class="bi bi-exclamation-circle text-xl"></i>
        All data on this page is strictly for demonstration purposes and fake.
    </p>
</div>
<div class="grid gap-y-12">

    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3 bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">peace gas top</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1987-06-04 </p>
                <p class="text-right">By William Stewart </p>
            </div>
            <p>Base impact heart few kitchen sound husband. Fall glass heart section star if.
Trip successful outside audience by. House subject past should.</p>
        </div>
    </div>

    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3 bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">my receive speech</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1990-03-29 </p>
                <p class="text-right">By Steven Moreno </p>
            </div>
            <p>Make education common. Without give talk time always. White ever you must soldier eye.
Particularly course ago bad near guess. Machine blood write mind young behavior. To really hear finally go.</p>
        </div>
    </div>

    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3 bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">good fact then</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1986-11-05 </p>
                <p class="text-right">By Christopher Mason </p>
            </div>
            <p>Rate discuss house back table. Base myself world young leader mother.
Son employee compare pick reality specific. Could course wrong race. Allow standard technology real maintain add.</p>
        </div>
    </div>

...
```

Notice the repeating structure

```
    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3
bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">peace gas top</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1987-06-04 </p>
                <p class="text-right">By William Stewart </p>
            </div>
            <p>Base impact heart few kitchen sound husband. Fall
glass heart section star if.
Trip successful outside audience by. House subject past should.</p>
        </div>
    </div>
```

```
    <div class="grid grid-cols-4 gap-x-4 border rounded pr-3
bg-green-50 hover:shadow-lg transition duration-500">
        <img src="/static/placeholder.png" />
        <div class="col-span-3 space-y-3 py-3">
            <h2 class="text-2xl text-green-900">my receive
speech</h2>
            <div class="grid grid-cols-2 italic">
                <p> 1990-03-29 </p>
                <p class="text-right">By Steven Moreno </p>
            </div>
            <p>Make education common. Without give talk time always.
White ever you must soldier eye.
Particularly course ago bad near guess. Machine blood write mind
young behavior. To really hear finally go.</p>
        </div>
    </div>
```

Structure is the same, element contents are different

Mini Exercise: Web Scraping People

1. Inspect the page at https://web-scraping-demo.zgulde.net/people. What classes and html structure can you use to extract the information you want?
2. Write python code to turn the page contents into a soup object.
3. Write the code necessary to loop through all of the people and extract their information.
4. The result should be a pandas DataFrame.