

# Installing BeautifulSoup

```
pip install beautifulsoup4
```

## Scrape Book Titles and Prices (30 pts)

Write a Python program that scrapes book information (title and price) from the website <https://books.toscrape.com/> and extracts all the books and their prices.

### Website Description

The website lists books in a simple HTML structure.

Each book entry includes:

- A `<h3>` tag containing the book title (inside an `<a>` tag).
- A `<p>` tag with the class `"price_color"` containing the book's price.

Example HTML snippet:

```
<article class="product_pod">
    ...
    <h3><a href="..." title="A Light in the Attic">A Light in the Attic</a></h3>
    <div class="product_price">
        <p class="price_color">£51.77</p>
        ...
    </div>
</article>
```

### Program Requirements

#### 1. Fetch the web page

Use the `requests` library to download the HTML content from `https://books.toscrape.com/`.

*Hint:* The website contains multiple pages, but for this assignment, you only need to scrape the **first page** (`https://books.toscrape.com/`).

#### 2. Parse the HTML

Use **BeautifulSoup** from the `bs4` library to parse the page and extract:

- The **title** of each book.

- The **price** of each book.

### 3. Store the data

Create a dictionary where:

- The **key** is the book title.
- The **value** is the book price as a **float** (for example, 51.77 instead of '£51.77' ).

### 4. Save the results to a file

Write the collected data to a text file named `books.txt` , with each line formatted as:

```
51.77, A Light in the Attic  
...
```

## Parse Country Names and Capitals (20 pts)

Write a Python program that fetches and parses country information (name and capital city) from the World Bank's XML API and extracts multiple countries' names and capitals.

### API Description

Use the following URL to get a list of countries in XML format:

```
https://api.worldbank.org/v2/country?per_page=50
```

Each `<wb:country>` element in the XML contains:

- `<wb:name>` — the country name
- `<wb:capitalCity>` — the capital city

Example snippet:

```
<wb:country id="BRA">  
  <wb:iso2Code>BR</wb:iso2Code>  
  <wb:name>Brazil</wb:name>  
  ...  
  <wb:capitalCity>Brasilia</wb:capitalCity>  
</wb:country>
```

# Program Requirements

## 1. Fetch the XML data

Use the `requests` library to download the XML content from `https://api.worldbank.org/v2/country?per_page=50` .

## 2. Parse the XML

Use Python's built-in `xml.etree.ElementTree` module (or `BeautifulSoup` with `parser="xml"` ) to extract:

- The **country name** from `<wb:name>`
- The **capital city** from `<wb:capitalCity>` — if the value is empty, treat it as "N/A"

## 3. Store the data

Create a dictionary where:

- The **key** is the country name.
- The **value** is the capital city.

## 4. Save the results to a file

Write the collected data to a text file named `countries.txt` , with each line formatted as:

```
Country: Brazil, Capital: Brasilia
Country: United States, Capital: Washington D.C.
Country: Antarctica, Capital: N/A
```