

Eric Heitman

Python Library: BeautifulSoup

Project Assignment #2: Write-Up Part 1

4/8/20

Python Library: BeautifulSoup

Background and Central Purpose:

A popular python library known as BeautifulSoup is helpful in providing a wealth of options for extracting data from web files. From HTML and XML documents, this web scraping tool leverages parsers to allow for easy navigation, modification, and searchability through web data, which ultimately results in a huge boost of efficiency for many programmers.

Practical Applications:

The realm of possibility of applications of BeautifulSoup are seemingly endless. During my research, I found many applications of BeautifulSoup in the realm of checking price changes to products on online marketplaces such as Ebay or Amazon. Programmers would write code to locate the specific div in which this information exists, and then extract this information and use it in some meaningful way. However, it is important to note that specific sites like Amazon block the ability to scrape data without the use of web API's.

Other interesting ways of using BeautifulSoup would be to write Python code to have up to date data on your own website. For example, you could write Python code and implement it via XML code to display the current weather forecast for specific location on the banner of your website. This could also be helpful if you run a website dedicated to identifying changes in weather over time in specific areas.

Overview of Functions & Further References:

The amount of functions BeautifulSoup contains are extensive. They provide programmers with the ability to parse information and navigate it effectively to find what you are

looking for, such as the `.string` function, which allows you to do a string comparison to find any data associated with the string you specify. Popular functions include `.find` (finds a specific div as a parameter to the function), `.get_text` (extracts the text part of a document tag), and `SoupStrainer` (a package which allows for easy searches for only specific parts of a web page, for example web links). For further reference, please visit the official documentation page for BeautifulSoup at <https://www.crummy.com/software/BeautifulSoup/bs4/doc/#> .