Alloy Abasali

Prof. Robert Domanski

CPS 3320 - 01

April 6th, 2022

BeautifulSoup library

Beautiful Soup is a Python library that is used for web scraping purposes to pull the data out of HTML and XML files. It creates a parse tree from the page source code, which can be used to extract data in a hierarchical and more readable manner. BeautifulSoup can be used for web Scraping used by companies to scrap the product data for their products and competing products as well to see how it impacts their pricing strategies. Companies also use this data to fix the optimal pricing for their products so that they can obtain maximum revenue. BeautifulSoup can also be used by companies for email marketing.

Overview of the different functions in the BeautifulSoup library

BeautifulSoup(): BeautifulSoup(,) creates a data structure representing a parsed HTML or XML document.

find(): the find() function is used to search for anything in our web page.

get\_text(): this function is used text extract the text from the elements.

strip(): The strip() method returns a copy of the string with both leading and trailing characters removed.

split(): This function splits the string into different parts and we can use the parts that we desire.

It works with a combination of the separator and a string.

Links:

https://pypi.org/project/beautifulsoup4/

https://www.crummy.com/software/BeautifulSoup/bs4/doc/