**WEB SCRAPING**

Web Scraping helps to extract data from websites to collect useful information from it. It sends a GET request to a scraping website to load the website or the webpage and then scraping tools are used to perform scraping. Web scraping is the process of using bots to extract content and data from a website. It extracts HTML and XML code.

**BEAUTIFUL SOUP**

Beautiful soup is a python module that is used in web scraping to parse HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML

|  |  |  |
| --- | --- | --- |
| **Things I’ve learned** | **Type** | **Meaning** |
| BeautifulSoup(page.text, ‘html.parser’) | class | Beautiful soup object that is used to parse html document. The first argument is the loaded webpage from server and second argument is used to format that webpage in html format, because by default, it loads the jungle of text. |
| soup.find(tag, object) | method | This method will find the first occurrence of any tag.  The first argument is which tag do you want to find and second argument is for uniqueness like what class it belongs. The second argument is optional.  soup.find(‘div’, {‘class’:’w3-bar-block’}) |
| soup.findAll(tag, object) | method | This method will find all the occurrence of any tag and store them in a list.  soup.findAll(‘div’, {‘class’:’w3-bar-block’}) |
| soup.findChild() | method | This method will find the first occurrence of any child tag in a parent tag.  soup.findChild(‘a’)  This is will find first occurrence of – a – tag in the object soup. |
| soup.findChildren() | method | This method will find all the occurrence of children tag in a parent tag.  soup.findChildren(‘a’)  This is will find all the occurrence of – a – tag in the object soup and store them in a list.  By default it takes zero arguments and can find all the children of the parent tag rather than going for specific tag. |