

Web Scrapping: A Key Tool in Data Science

Estimated Effort: 5 mins

Introduction

Web scrapping, also known as web harvesting or web data extraction, is a technique used to extract large amounts of data from websites. The data on websites is unstructured, and web scrapping enables us to convert it into a structured form.

Importance of Web Scrapping in Data Science

In the field of data science, web scrapping plays an integral role. It is used for various purposes such as:

- 1. **Data Collection:** Web scrapping is a primary method of collecting data from the internet. This data can be used for analysis, research, etc.
- 2. **Real-time Applications:** Web scrapping is used for real-time applications like weather updates, price comparison, etc.
- 3. **Machine Learning:** Web scrapping provides the data needed to train machine learning models.

Web Scrapping with Python

Python provides several libraries for web scrapping. Here are some of them:

- 1. **BeautifulSoup:** BeautifulSoup is a Python library used for web scrapping purposes to pull the data out of HTML and XML files. It creates a parse tree from page source code that can be used to extract data in a hierarchical and more readable manner.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
9. 9
10. 10
11. 11
12. 12
13. 13
14. 14
15. 15
16. 16
17. 17
18. 18
19. 19
20. 20
21. 21
22. 22
23. 23
24. 24
25. 25
26. 26
27. 27
28. 28
29. 29
30. 30
31. 31
32. 32
33. 33
34. 34
35. 35
36. 36
37. 37
38. 38
39. 39
40. 40
41. 41
42. 42
43. 43
44. 44
45. 45
46. 46
47. 47
48. 48
49. 49
50. 50
51. 51
52. 52
53. 53
54. 54
55. 55
56. 56
57. 57
58. 58
59. 59
60. 60
61. 61
62. 62
63. 63
64. 64
65. 65
66. 66
67. 67
68. 68
69. 69
70. 70
71. 71
72. 72
73. 73
74. 74
75. 75
76. 76
77. 77
78. 78
79. 79
80. 80
81. 81
82. 82
83. 83
84. 84
85. 85
86. 86
87. 87
88. 88
89. 89
90. 90
91. 91
92. 92
93. 93
94. 94
95. 95
96. 96
97. 97
98. 98
99. 99
100. 100
```

- 2. **Scrapy:** Scrapy is an open-source and collaborative web crawling framework for Python. It is used to extract the data from the website.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
9. 9
10. 10
11. 11
12. 12
13. 13
14. 14
15. 15
16. 16
17. 17
18. 18
19. 19
20. 20
21. 21
22. 22
23. 23
24. 24
25. 25
26. 26
27. 27
28. 28
29. 29
30. 30
31. 31
32. 32
33. 33
34. 34
35. 35
36. 36
37. 37
38. 38
39. 39
40. 40
41. 41
42. 42
43. 43
44. 44
45. 45
46. 46
47. 47
48. 48
49. 49
50. 50
51. 51
52. 52
53. 53
54. 54
55. 55
56. 56
57. 57
58. 58
59. 59
60. 60
61. 61
62. 62
63. 63
64. 64
65. 65
66. 66
67. 67
68. 68
69. 69
70. 70
71. 71
72. 72
73. 73
74. 74
75. 75
76. 76
77. 77
78. 78
79. 79
80. 80
81. 81
82. 82
83. 83
84. 84
85. 85
86. 86
87. 87
88. 88
89. 89
90. 90
91. 91
92. 92
93. 93
94. 94
95. 95
96. 96
97. 97
98. 98
99. 99
100. 100
```

- 3. **Selenium:** Selenium is a tool used for controlling web browsers through programs and automating browser tasks.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
9. 9
10. 10
11. 11
12. 12
13. 13
14. 14
15. 15
16. 16
17. 17
18. 18
19. 19
20. 20
21. 21
22. 22
23. 23
24. 24
25. 25
26. 26
27. 27
28. 28
29. 29
30. 30
31. 31
32. 32
33. 33
34. 34
35. 35
36. 36
37. 37
38. 38
39. 39
40. 40
41. 41
42. 42
43. 43
44. 44
45. 45
46. 46
47. 47
48. 48
49. 49
50. 50
51. 51
52. 52
53. 53
54. 54
55. 55
56. 56
57. 57
58. 58
59. 59
60. 60
61. 61
62. 62
63. 63
64. 64
65. 65
66. 66
67. 67
68. 68
69. 69
70. 70
71. 71
72. 72
73. 73
74. 74
75. 75
76. 76
77. 77
78. 78
79. 79
80. 80
81. 81
82. 82
83. 83
84. 84
85. 85
86. 86
87. 87
88. 88
89. 89
90. 90
91. 91
92. 92
93. 93
94. 94
95. 95
96. 96
97. 97
98. 98
99. 99
100. 100
```

Applications of Web Scrapping

Web scrapping is used in various fields and has many applications:

- 1. **Price Comparison:** Services such as Parallels use web scrapping to collect data from online shopping websites and use it to compare the prices of products.
- 2. **Email address gathering:** Many companies that use email as a medium for marketing, use web scrapping to collect email ID and then send bulk emails.
- 3. **Social Media Scrapping:** Web scrapping is used to collect data from Social Media websites such as Twitter to find out what's trending.

Conclusion

Web scrapping is an essential skill in the fast-growing world of data science. It provides the ability to turn the web into a source of data that can be analyzed, processed, and used for a variety of applications. However, it's important to remember that one should use web scrapping responsibly and ethically, respecting the terms of use or robots.txt files of the websites being scrapped.

Author(s)

[Abhishek Gargula](#)

