

Web Scraping — Practice

CSCI 40: Computing for the Web

Fall 2026

Quiz rules:

1. You MAY use any printed or handwritten notes.
2. You MAY NOT use a computer or any other electronic device.

For each BeautifulSoup selector below, write how many tags it matches (the graded quiz follows this pattern).

quiz_practice_bs4.py

```
1 '''
2 Beautiful Soup Practice Quiz
3 =====
4 '''
5
6 from bs4 import BeautifulSoup
7 html'''
8 <html>
9     <head>
10         <title>Lorem Ipsum</title>
11     </head>
12     <body>
13         <div id=header>
14             <img src=top-image.png>
15         </div>
16         <h1 id=top>Lorem <span class=bold>Ipsum</span></h1>
17         <p><span class=italic>Lorem ipsum</span> dolor sit amet, consectetur adipiscing elit, sed
18             do eiusmod tempor incididunt ut labore et dolore magna aliqua.</p>
19         <blockquote><span class=bold>Ut enim</span> ad minim veniam, quis nostrud exercitation
20             ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in
21             reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.</
22             blockquote>
23         <p class=bold><a href=https://izbicki.me>Excepteur sint occaecat</a> cupidatat non proident
24             , sunt in culpa qui officia deserunt mollit anim id est laborum.</p>
25         <br/>
26         <a href=https://izbicki.me><img src=bottom-image.png></a>
27     </body>
28 </html>
29 '''
30
31 soup = BeautifulSoup(html, 'html.parser')
32
33 #####
34 tags = soup.select('a')
35 problem1 = len(tags)
36 print("problem1=",problem1)
37
38 #####
39 tags = soup.select('p span')
40 problem2 = len(tags)
41 print("problem2=",problem2)
42
43 #####
44 tags = soup.select('p,.bold') # remove duplicates
45 problem3 = len(tags)
46 print("problem3=",problem3)
```

```

47
48
49 #####
50 # equivalent to 'p a'
51 tags = soup.select('p')
52 accumulator = []
53 for tag in tags:
54     accumulator += tag.select('a')
55 problem4 = len(accumulator)
56 print("problem4=",problem4)
57
58
59
60 #####
61 tags1 = soup.select('p') # do not remove duplicates
62 tags2 = soup.select('.bold')
63 tags = tags1 + tags2
64 problem5 = len(tags)
65 print("problem5=",problem5)

```

quiz_practice_bs4_2.py

```

1  '''
2  Beautiful Soup Practice Quiz 2
3  =====
4  '''
5
6  from bs4 import BeautifulSoup
7  html'''
8  <html>
9      <head>
10         <title>Lorem Ipsum</title>
11      </head>
12      <body>
13         <div id=header>
14             <h1 id=top>Lorem <span class=bold>Ipsum</span></h1>
15             <img src=top-image.png>
16         </div>
17         <div id=body>
18             <h2>Lorem</h2>
19             <p class=bold>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod
20                 tempor incididunt ut labore et dolore magna aliqua.</p>
21             <p class=bold>Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut
22                 aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in
23                 voluptate velit esse cillum dolore eu fugiat nulla pariatur.</p>
24             <h2>Ipsum</h2>
25             <ol>
26                 <li><img class=bold src=ceasar.png></li>
27                 <li class=bold><em>Veni</em></li>
28                 <li class=bold><em>Vidi</em></li>
29                 <li class=bold><em>Vici</em></li>
30             </ol>
31         </div>
32     </body>
33 </html>
34 '''
35
36 soup = BeautifulSoup(html, 'html.parser')
37
38 #####
39 tags = soup.select('div > li')
40 problem1 = len(tags)
41 print("problem1=",problem1)
42
43 #####
44 tags = soup.select('div li')
45 problem1a = len(tags)

```

```

43 print("problem1a=",problem1a)
44
45 #####
46 tags = soup.select('img.bold')
47 problem1b = len(tags)
48 print("problem1b=",problem1b)
49
50 #####
51 tags = soup.select('li.bold')
52 problem1c = len(tags)
53 print("problem1c=",problem1c)
54
55 #####
56 tags = soup.select('li .bold')
57 problem1d = len(tags)
58 print("problem1d=",problem1d)
59
60 #####
61 tags = soup.select('h2+p')
62 problem1e = len(tags)
63 print("problem1e=",problem1e)
64
65 #####
66 tags = soup.select('div>h2+p,h2')
67 problem1f = len(tags)
68 print("problem1f=",problem1f)
69
70 #####
71 tags = soup.select('ol')
72 accumulator = []
73 for tag in tags:
74     accumulator += tag.select('img')
75 problem2 = len(accumulator)
76 print("problem2=",problem2)
77
78 #####
79 tags = soup.select('div')
80 accumulator = []
81 for tag in tags:
82     accumulator += tag.select('img')
83 problem2a = len(accumulator)
84 print("problem2a=",problem2a)
85
86 #####
87 tags = soup.select('body')
88 accumulator = []
89 for tag in tags:
90     accumulator += tag.select('title')
91 problem2b = len(accumulator)
92 print("problem2b=",problem2b)
93
94 #####
95 tags = soup.select('html')
96 accumulator = []
97 for tag in tags:
98     accumulator += tag.select('title')
99 problem2c = len(accumulator)
100 print("problem2c=",problem2c)
101
102 #####
103 tags1 = soup.select('h1')
104 tags2 = soup.select('#top')
105 tags = tags1 + tags2
106 problem3 = len(tags)
107 print("problem3=",problem3)
108
109 #####
110 tags1 = soup.select('p')

```

```
111 tags2 = soup.select('.bold')
112 tags = tags1 + tags2
113 problem3b = len(tags)
114 print("problem3b=",problem3b)
115
116
117 #####
118 tags1 = soup.select('div')
119 tags2 = soup.select('.bold')
120 tags = tags1 + tags2
121 problem3c = len(tags)
122 print("problem3c=",problem3c)
```