

Web Scraping with BeautifulSoup — Answer Key

CSCI 40: Computing for the Web

Fall 2026

```
1 <html>
2   <head>
3     <title>Lorem Ipsum</title>
4   </head>
5   <body>
6     <div id=header>
7       <img src=top-image.png>
8     </div>
9     <h1 id=top>Lorem <span class=bold>Ipsum</span></h1>
10    <p><span class=italic>Lorem ipsum</span> dolor sit amet, consectetur adipiscing el
11    <blockquote><span class=bold>Ut enim</span> ad minim veniam, quis nostrud exercitat
12    <p class=bold><a href=https://izbicki.me>Excepteur sint occaecat</a> cupidatat non
13    <br/>
14    <a href=https://izbicki.me><img src=bottom-image.png></a>
15  </body>
16 </html>
```

Problem 1. What value does `problem1` hold after this code runs?

```
1 tags = soup.select('a')
2 problem1 = len(tags)
```

Answer:

2

Problem 2. What value does `problem2` hold after this code runs?

```
1 tags = soup.select('p_span')
2 problem2 = len(tags)
```

Answer:

1

Problem 3. What value does `problem3` hold after this code runs?

```
1 tags = soup.select('p,.bold') # remove duplicates
2 problem3 = len(tags)
```

Answer:

4

Problem 4. What value does problem4 hold after this code runs?

```
1 # equivalent to 'p a'
2 tags = soup.select('p')
3 accumulator = []
4 for tag in tags:
5     accumulator += tag.select('a')
6 problem4 = len(accumulator)
```

Answer:

1

Problem 5. What value does problem5 hold after this code runs?

```
1 tags1 = soup.select('p') # do not remove duplicates
2 tags2 = soup.select('.bold')
3 tags = tags1 + tags2
4 problem5 = len(tags)
```

Answer:

5

```
1 <html>
2     <head>
3         <title>Lorem Ipsum</title>
4     </head>
5     <body>
6         <div id=header>
7             <h1 id=top>Lorem <span class=bold>Ipsum</span></h1>
8             <img src=top-image.png>
9         </div>
10        <div id=body>
11            <h2>Lorem</h2>
12            <p class=bold>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do e
13            <p class=bold>Ut enim ad minim veniam, quis nostrud exercitation ullamco labor
14            <h2>Ipsum</h2>
15            <ol>
16                <li><img class=bold src=ceasar.png></li>
17                <li class=bold><em>Veni</em></li>
18                <li class=bold><em>Vidi</em></li>
19                <li class=bold><em>Vici</em></li>
20            </ol>
21        </div>
22    </body>
23 </html>
```

Problem 6. What value does `problem1` hold after this code runs?

```
1 tags = soup.select('div>li')
2 problem1 = len(tags)
```

Answer:

0

Problem 7. What value does `problem1a` hold after this code runs?

```
1 tags = soup.select('div_li')
2 problem1a = len(tags)
```

Answer:

4

Problem 8. What value does `problem1b` hold after this code runs?

```
1 tags = soup.select('img.bold')
2 problem1b = len(tags)
```

Answer:

1

Problem 9. What value does `problem1c` hold after this code runs?

```
1 tags = soup.select('li.bold')
2 problem1c = len(tags)
```

Answer:

3

Problem 10. What value does `problem1d` hold after this code runs?

```
1 tags = soup.select('li_.bold')
2 problem1d = len(tags)
```

Answer:

1

Problem 11. What value does `problem1e` hold after this code runs?

```
1 tags = soup.select('h2+p')
2 problem1e = len(tags)
```

Answer:

1

Problem 12. What value does `problem1f` hold after this code runs?

```
1 tags = soup.select('div>h2+p,h2')
2 problem1f = len(tags)
```

Answer:

3

Problem 13. What value does `problem2` hold after this code runs?

```
1 tags = soup.select('ol')
2 accumulator = []
3 for tag in tags:
4     accumulator += tag.select('img')
5 problem2 = len(accumulator)
```

Answer:

1

Problem 14. What value does `problem2a` hold after this code runs?

```
1 tags = soup.select('div')
2 accumulator = []
3 for tag in tags:
4     accumulator += tag.select('img')
5 problem2a = len(accumulator)
```

Answer:

2

Problem 15. What value does `problem2b` hold after this code runs?

```
1 tags = soup.select('body')
2 accumulator = []
3 for tag in tags:
4     accumulator += tag.select('title')
5 problem2b = len(accumulator)
```

Answer:

0

Problem 16. What value does `problem2c` hold after this code runs?

```
1 tags = soup.select('html')
2 accumulator = []
3 for tag in tags:
4     accumulator += tag.select('title')
5 problem2c = len(accumulator)
```

Answer:

1

Problem 17. What value does `problem3` hold after this code runs?

```
1 tags1 = soup.select('h1')
2 tags2 = soup.select('#top')
3 tags = tags1 + tags2
4 problem3 = len(tags)
```

Answer:

2

Problem 18. What value does `problem3b` hold after this code runs?

```
1 tags1 = soup.select('p')
2 tags2 = soup.select('.bold')
3 tags = tags1 + tags2
4 problem3b = len(tags)
```

Answer:

9

Problem 19. What value does `problem3c` hold after this code runs?

```
1 tags1 = soup.select('div')
2 tags2 = soup.select('.bold')
3 tags = tags1 + tags2
4 problem3c = len(tags)
```

Answer:

9