

master ...[html2csv](#) / [html2csv](#) / [converter.py](#) / [Code](#) Jump to ▼ **jumbroso** Fixed typo with table name reused History 2 contributors52 lines (43 sloc) | 1.38 KB ...

```
1 import csv
2 import io
3
4 import bs4
5
6
7 def detect_engine():
8     try:
9         import lxml
10    except ImportError:
11        engine = 'html.parser'
12    else:
13        engine = 'lxml'
14    return engine
15
16
17 class Converter:
18
19     def __init__(self, **kwargs):
20         engine = kwargs.get('engine')
21         if engine is None:
22             self.engine = detect_engine()
23         else:
24             self.engine = engine
25         self.params = kwargs
26
27     def convert_to_list(self, html_doc):
28         soup = bs4.BeautifulSoup(html_doc, self.engine)
29         tables = []
30         for table_element in soup.find_all('table'):
31             table = []
32             for tr in table_element.find_all('tr'):
33                 row = [''.join(cell.stripped_strings)
34                       for cell in tr.find_all(['td', 'th'])]
35                 table.append(row)
36             tables.append(table)
37         return tables
38
39     def convert_to_csv(self, html_doc):
40         tables = self.convert_to_list(html_doc)
41         output = []
42         for table_num, table in enumerate(tables):
43             csv_string = io.StringIO()
44             csv_writer = csv.writer(csv_string)
45             for row in table:
46                 csv_writer.writerow(row)
47             table_attrs = dict(num=table_num)
48             output.append((csv_string.getvalue(), table_attrs))
49         return output
50
51     def convert(self, html_doc):
52         return self.convert_to_csv(html_doc)
```