

Forum:[Hack Forums](#)

Thread: [AutoScrapper automatically deletes duplicates... how to keep them?](#)

essie98927 - April 22, 2022, 1:05 pm

I'm trying to use the AutoScrapper module with python 3.9.6 to scrape some information on websites but it deletes automatically duplicate information but i need them...

****I'm trying to scrape the % in the March column**

Here's my code :

```
Code import autoscraper
from autoscraper import AutoScrapper
```

```
url = "https://store.steampowered.com/hwsurvey/videocard/"
```

```
wanted_list = ["8.18%"]
```

```
scraper = AutoScrapper()
result_per100 = scraper.build(url, wanted_list)
```

```
print(result_per100)
```

Output:

```
Code ['92.00%', '1.94%', '1.12%', '0.03%', '0.00%', '4.91%', '8.18%', '6.08%', '5.62%', '5.53%',
'2.87%', '2.71%', '2.48%', '2.45%', '2.35%', '2.22%', '2.11%', '2.06%', '1.98%', '1.65%', '1.55%',
'1.54%', '1.46%', '1.41%', '1.32%', '1.27%', '1.25%', '1.19%', '1.11%', '1.08%', '1.04%', '1.02%',
'0.95%', '0.94%', '0.89%', '0.86%', '0.83%', '0.77%', '0.72%', '0.70%', '0.69%', '0.68%', '0.64%',
'0.63%', '0.56%', '0.54%', '0.51%', '0.50%', '0.47%', '0.46%', '0.45%', '0.44%', '0.42%', '0.39%',
'0.37%', '0.36%', '0.35%', '0.34%', '0.33%', '0.30%', '0.27%', '0.26%', '0.25%', '0.24%', '0.23%',
'0.22%', '0.21%', '0.19%', '0.18%', '0.17%', '0.16%', '10.02%', '8.64%', '6.40%', '5.93%', '5.83%',
'3.02%', '2.86%', '2.61%', '2.58%', '2.49%', '2.33%', '2.17%', '2.09%', '1.74%', '1.63%', '1.62%',
'1.48%', '1.39%', '1.34%', '1.31%', '1.16%', '1.15%', '1.10%', '1.00%', '0.99%', '0.93%', '0.91%',
'0.87%', '0.81%', '0.76%', '0.66%', '0.59%', '0.57%', '0.53%', '0.48%', '0.41%', '0.38%', '0.31%',
'14.83%', '24.09%', '17.74%', '15.10%', '6.56%', '6.15%', '2.93%', '2.24%', '2.20%', '2.10%', '1.84%',
'1.73%', '1.14%', '1.07%', '0.79%', '0.74%', '0.73%', '0.65%', '0.58%', '0.52%', '0.43%', '0.32%',
'1.58%', '24.30%', '14.36%', '6.70%', '6.32%', '2.84%', '2.26%', '2.02%', '2.00%', '1.69%', '0.92%',
'0.84%', '0.62%', '0.60%', '0.55%', '10.69%', '22.32%', '14.34%', '13.79%', '7.08%', '5.08%', '2.90%',
'2.72%', '2.36%', '1.81%', '1.45%', '1.09%', '6.72%', '21.62%', '16.22%', '10.81%', '8.11%', '5.41%',
'2.70%']
```

as you can see there's no duplicates...

Anyone knows if that can be fixed and how ?

Any help is much appreciated thanks !

Don Cezar - April 22, 2022, 1:52 pm

As someone commented on your stackoverflow thread, we need more information and your code to figure out what's going on

essie98927 - April 22, 2022, 2:04 pm

(Apr 22, 2022 01:52 PM)goldieczr Wrote: [\[CITE\]](#) As someone commented on your stackoverflow thread, we need more information and your code to figure out what's going on
thought it was a bot sorry

I don't know what to add more actually I'll try to make a dummy code and post it but don't know if that will help cause it seems that the module does that by itself as default

(Apr 22, 2022 01:52 PM)goldieczr Wrote: [\[CITE\]](#) As someone commented on your stackoverflow thread, we need more information and your code to figure out what's going on
fixed it ;)

StandForFreedom - April 24, 2022, 12:20 am

It's because your using Autoscraper.Build method instead of get the get results method.

The other methods like `get_result_exact`, `get_result` have a parameter called "unique" that you can set to false to keep duplicates.

The Build method in the code does not have that parameter but in the code does remove duplicates as seen on line 253 in `auto_scraper.py` calling `unique_hashable` method.

```
Code result_list = [item.text for item in result_list]
      result_list = unique_hashable(result_list)
      self.stack_list = unique_stack_list(self.stack_list)
That method executes this code in utils.py on line 22
```

```
Code def unique_hashable(hashable_items):
    """Removes duplicates from the list. Must preserve the orders."""
    return list(OrderedDict.fromkeys(hashable_items))
```

So your option is to use the other methods or modify the library you are using to not call on these methods.

essie98927 - April 25, 2022, 2:40 am

(Apr 24, 2022 12:20 AM)StandForFreedom Wrote: [\[CITE\]](#) It's because your using Autoscraper.Build method instead of get the get results method.

The other methods like `get_result_exact`, `get_result` have a parameter called "unique" that you can set to false to keep duplicates.

The Build method in the code does not have that parameter but in the code does remove duplicates as seen on line 253 in `auto_scraper.py` calling `unique_hashable` method.

```
Code result_list = [item.text for item in result_list]
      result_list = unique_hashable(result_list)
      self.stack_list = unique_stack_list(self.stack_list)
That method executes this code in utils.py on line 22
```

```
Code def unique_hashable(hashable_items):
    """Removes duplicates from the list. Must preserve the orders."""
```

```
return list(OrderedDict.fromkeys(hashable_items))
```

So your option is to use the other methods or modify the library you are using to not call on these methods.

thank you very much for you detailed answer I'll try those different options !