

Forum:[Hack Forums](#)

Thread: [Remove duplicate lines from txt file? QUICK](#)

Woes - April 22, 2017, 7:24 pm

I googled this, seems simple af, but i'm trying some of the code and none of it is working, and I don't really feel like thinking it out.

Can somebody shoot me some code that works. Thanks!

(???)?(????) - April 22, 2017, 7:53 pm

<http://stackoverflow.com/questions/1215208/how-might-i-remove-duplicate-lines-from-a-file>

PrivateDonut - April 22, 2017, 7:54 pm

Do you need the actual code? Because I know of a website that will remove duplicate lines.

CaptainBlack - April 22, 2017, 8:53 pm

PHP Code

```
<?php
def rm_dup_lines(file_path):
    with open(file_path) as f:
        contents = f.readlines()
    for line in contents:
        while contents.count(line) > 1:
            contents.remove(line)
    with open (file_path, "w") as f:
        f.write("".join(contents))
```

Or simply use a set, sets in Python don't allow duplications:

PHP Code

```
<?php
def rm_dup_lines(file_path):
    with open(file_path) as f:
        contents = f.readlines()
    result = set(contents)
    with open(file_path) as f:
        f.write("".join(result))
```

systx - April 23, 2017, 11:38 am

Quick and easy way that works perfectly...

PHP Code

```
<?php
read = open("FILE NAME", 'r').readlines()
read_set = set(read)
new_path = NEW NAME
out = open(new_path, 'w')
for line in read_set:
```

```
out.write(line)
```

SIM-1866423 - April 23, 2017, 11:53 am

Well everyone seems like agree on same thing use set function coz its built-in function.

.kr. - April 23, 2017, 1:39 pm

```
Code #!/usr/bin/python
import sys
```

```
def main(argc, argv):
    print "".join(set(open(argv[1]).readlines())) if argc == 2 else "Usage\n {} filename".format(argv[0])
    return 0

if __name__ == "__main__":
    main(len(sys.argv), sys.argv)
```

TheWarOwl - April 23, 2017, 2:32 pm

Here's a simple solution - store all of the new lines into a list and use set:

PHP Code

```
<?php
f = open("file.txt","r") #Open the file for reading
data = f.readlines() #Creates the list for each line
newfile = set(data) #Set removes duplicates in a list
f.close() #Close the file for later use
```

Then you can simply write over the old file or create a new one!

PHP Code

```
<?php
f = open("file.txt","w") #Opens the file for writing - will write over old data
f.write(newfile) #Here we write the information we want
f.close() #Once again - close the file
```

There are more simple solutions and one liners, however I want you to understand the process the best you can so I've expanded it for you. ?

RyanR45 - May 1, 2017, 6:33 pm

(Apr 22, 2017 07:24 PM)Woes Wrote: [\[CITE\]](#) I googled this, seems simple af, but i'm trying some of the code and none of it is working, and I don't really feel like thinking it out.

Can somebody shoot me some code that works. Thanks!

Hey can you PM me please I have a question about your services

Goodies - May 1, 2017, 8:35 pm

This would be a way to do it with the least amount of code. It also assumes that order is not important.

PHP Code

```
<?php
def unique_lines(filename):
    with open(filename, "r") as f:
```

```
return list(map(str.strip, set(f.readlines()))))
```

Test file:

Code this is a test

testing 123

this is a test

testing 123

this is not a test

Code >>> unique_lines("lol.txt")

['this is not a test', 'testing 123', 'this is a test']