

Forum:[Hack Forums](#)

Thread: [\[\\$100\] I need to program to scan my external drive for duplicate files](#)

Reporter - April 20, 2017, 8:37 am

I need a GOOD program to scan my computer (External USB drive) to find duplicates (Such as photos, music, videos ect.).

The program need to make sure that it is a correct 100% same file and not 2 different files in different folder but with the same name.

The program can have a software license that can cost up to €100+.

Can anyone recommend a Free / Paid software that can do this?

mistakes should not happen.

CaptainBlack - April 20, 2017, 6:23 pm

Here's an open source Python script written by me that does exactly that:

<https://github.com/MonroCoury/PySpace>

Description is in the readme.md

CaptainBlack - April 20, 2017, 6:55 pm

^First of all, big thanks for the rep!

I have researched the topic actually, and it turns out that the odds of 2 different files producing the same md5 hash code are practically negligibly small:

<https://crypto.stackexchange.com/questions/18335/probability-that-two-different-files-can-have-same-md5-hash-for-both-an-unencryp>

<https://superuser.com/questions/556923/whats-the-percentage-of-having-two-files-with-the-same-byte-size-length-giving>

SHA-256 would lead to even more unlikely collisions, but the reason why I used md5 is because generating md5 hashes takes less time and resources.

Tho it is "theoretically" possible.. So I'll update the readme.

Reporter - April 21, 2017, 1:13 am

(Apr 20, 2017 06:23 PM)CaptainBlack Wrote: [\[CITE\]](#) Here's an open source Python script written by me that does exactly that:

<https://github.com/MonroCoury/PySpace>

Description is in the readme.md

Does the bot autoclean or Will i be able to see what files are duplicates?

aa11 - April 21, 2017, 1:16 am

CCleaner has duplicate finder, might be worth looking at.

CaptainBlack - April 21, 2017, 1:34 am

(Apr 21, 2017 01:13 AM)Reporter Wrote: [\[CITE\]](#) Does the bot autoclean or Will i be able to see what files are duplicates?

It autocleans.. But you've brought up an interesting suggestion.. I'll add that soon.

Reporter - April 21, 2017, 12:01 pm

(Apr 21, 2017 01:34 AM)CaptainBlack Wrote: [\[CITE\]](#) It autocleans.. But you've brought up an interesting suggestion.. I'll add that soon.

PM me or post in thread once it lets me see what files it will remove so I can double check.
?

CaptainBlack - April 22, 2017, 12:15 am

Done! The script will now display the full path of all the files it will delete and ask you for confirmation. If you enter anything that isn't y or yes it will abort.