

Find duplicate files

Scenario

Your system was breached and the attacker duplicated a lot of your files to random places, all over your drive.

Requirements

Write a program that returns a list of all the duplicate files. You can make the assumption that each file was duplicated only once.

A small accuracy hit is acceptable even if it means a small number of false positives.

The program will be used to help a user confirm that two files are indeed duplicates, and the user makes the deletion manually, since deleting files programmatically is really scary 🦴

The output of the program should be an array of file paths with the original and duplicate path.

For example:

```
(original: '/home/user/IMG_042.jpg', duplicate: '/var/log/nginx/access.log')  
(original: '/etc/nginx/nginx.conf', duplicate: '/home/suchsecurity.mp4')
```

- Since there is no single optimal solution, please present the benefits and tradeoffs of different approaches.
- Implement one of the presented approaches in the language of your choice.
- What is the space and time complexity of the solution? Can you bring it to $O(n)$ time and space, where n is the number of files?