CMPE 230 Systems Programming

Homework 2 (due May 12th)

Write a Python program that can be invoked with the following options and arguments:

```
identic  [-f | -d ] [-c] [-n] [-s] [<dir1> < dir2> ..]
```

The **identic** program will traverse the directories and look for files or directories that are duplicates of each other (i.e. identical). The full pathnames of duplicates will be printed as output (a new line should be printed between the sets of duplicates).

`[-f | -d ]`	-f means look for identical files, -d means look for identical directories. The default is identical files.
`-c`	Identical will mean the contents are exactly the same (note that the names can be different).
`-n`	Identical will mean the directory/file names are exactly the same (note that the contents can be different).
`-cn`	Identical will mean both the contents and the directory/file names are exactly the same.
`[<dir1> <dir2> ..]`	The list of directories to traverse (note that the directories will be traversed recursively, i.e. directories and their subdirectories and their subdirectories etc. etc.). The default is current directory.
`-s`	The size for each duplicate will also be printed. The duplicates should be printed in descending order of size. This option is to be ignored when –n option is used.

Note:

- Assume that directory hierarchy forms a tree. You can assume there will be no symbolic links.
- Use sha256 hashes in order to locate identical items.
- To locate identical directories, you can use hash trees.
- Use Python argparse package to parse the command line arguments: https://docs.python.org/3/library/argparse.html

Grading

Your project will be graded according to the following criteria:

Documentation (written document describing how you implemented your project)	12%
Comments in your code	8%
Implementation and tests	80%

Late Submission

If the project is submitted late, the following penalties will be applied:

- 0 < hours late <= 24 : 25%
- 24 < hours late <= 48 : 50%
- hours late > 48 : 100%

Timestamping

Project file should include your names in it. Please timestamp your project file using https://opentimestamps.org/ before you submit it. Keep the project file and its corresponding timestamp .ots file.