

CMPE230 HW#2

Objective

The duplicates program will traverse the directories and look for files or directories that are exact duplicates of each other. It will then carry out an action on the files or directories in question.

Solution

Format: duplicates < -c <command> | -p > [-f | -d] ["..."] [<dir1> <dir2> ..]

I used python3, and some libraries like os for directories and such, argsparse for parsing the arguments given in shell, hashlib for hashing the files. I used dictionaries a lot, specifically using a dictionary with format of {<hash_value>: [directories]} has been a great solution for finding duplicates.

Project Outline

- file_hasher(filepath): hashes the file.
 - dir_hasher(filepath): hashes directory.
 - find_dups(d_list, regexx): finds duplicates using directory list and regex for pattern matching.
 - command_execute(command,list): executes the given command to the directory list one by one. (Note: list is a list of lists.)
 - argument parser section: parses the arguments.
 - main loop: traverses all directories and hashes all the files and directories. Then appends all to a dictionary.
 - executing commands: if -f is given files are used, if not, directories are used. Uses command_execute() function.
-