Homework 3

16/09/2021

LINK: GitHub

1 Identify the names and format of the 3 biggest files. Can you come up with a command to generate a numerically ordered list of 3 biggest files? (hint: try using the wc command)

- changed working directory to folder where I have currently stored data (photos)
- `ls -F`
- listed content of the folder to examine the names and file format
- contains photos with serial number and various names describing content
- formats used are JPEG and RW2
- `wc -c *.JPG *.RW2 | sort -n | tail -n 4 | head -n 3`
- counts bytes in each file
- sorts by number
- takes 4 bottom filenames because the bottom numbers are biggest
- only shows 3 of the beginning numbers since wc includes the last line with total count

results:

1. 14713856 9237_Overview_W.RW2
2. 14733312 9247_Overview_SW.RW2
3. 14761472 9240_Overview_S.RW2

- test: opening file explorer and listing files by size we confirm these 3 to be the largest

2 Some of the image files are empty, a sign of error in the data processing or corruption. Generate a list of empty photo filenames to facilitate their later replacement easier? (hint: look up find)

- `find . -empty \( -name "*.JPG" -o -name "*.RW2" \) > list_empty_files.txt`
- finds all empty files in current directory ("HW" file) of which names include .JPG or .RW2 and assigns all names to a new .txt list
- result: list of all empty files, names displayed in format ./9260_Detail_Drain.JPG
- `cat list_empty_files.txt | cut -c 3- > list_empty_files.txt`
- removes the ./ characters from names of files

- `wc -l list_empty_files.txt`
- lists all lines in the list of empty files
- result: 73 list_empty_files.txt
- because I know each line corresponds to one empty file, I now know there is total of 73 empty photos which can be useful for later testing

3 **CHALLENGE (Easy to Medium): Write a shell script called emptyfiles.sh that takes the name of a directory and generates a list of empty photo filenames (remember to pick up all extensions). So that, for example, running bash emptyfiles.sh Downloads/badphotos produces a list of empty file filenames either in the terminal or in an external textfile. Can you eliminate the ./ notation at the start of the filename?**

- `nano emptyfiles.sh`
- opens nano editor, opens/creates shell script
- script:

```
for file in $@
do
find -empty \( -name "*.JPG" -o -name "*.RW2" \) | cut -c 3- > list_empty_files.txt
done
```

- similar script to 2nd task however adjusted so it can be used by another person on different data by specifying files when executing
- TESTING

```
bash emptyfiles.sh  /OneDrive/Desktop/w3/HW
```

- runs script on selected data

```
wc -l list_empty_files.txt
```

- output: 73 list_empty_files.txt
- same number of files as with previous code which suggests it worked

4 **CHALLENGE (Advanced): Imagine you have a directory goodphotos/ (same password as above) with original non-zero-length files sitting at the same level as the current directory. How would you write a loop to replace the zero-length files?**

- `nano replace.sh`
- opens editor and new script
- script:

```
for file in  /OneDrive/Desktop/w3/HW/*.JPG
do
filename=$(basename $file)
if grep -Fxql "$filename"  /OneDrive/Desktop/w3/HW/list_empty_files.txt;
then cp -f  /OneDrive/Desktop/w3/goodphotos/"$filename"  /OneDrive/Desktop/w3/HW
else echo "$filename"; fi;
done
```

- comparing filenames with the list of empty files
- if the filename matches, the empty file is replaced in HW
- if not, the name is printed out; we can use this to determine if the result is correct
- `bash replace.sh`
- runs script
- `find . -empty`
- result: no empty files are present in HW meaning they have all been replaced