5:W35: Managing Files on Steroids with Shell

DESCRIPTION

Your supervisor has shared a folder of photos on Sciencedata.dk with you (password is 2020CDS, folder is 500Mb and contains 189 images) and needs your help with a couple diagnostics:

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: consider using wc to gauge image size)

After downloading the folder with the 189 images, I managed it through the terminal. I went to the directory hw5_photos which I have named the folder with the images.
I used the command `wc -c *.JPG` to identify the size of the files in the directory. The size is shown in byte and it is the number written to the right before the file name. `wc -c` gives me the size and `*.JPG` matches all file names ending with ".JPG". All the file names ended on ".JPG" because jpg was the format of all of them.

```
[Emma-Maries-Air:hw5_photos Emma-Marie$ wc -c *.JPG
 1427378 9200_Overview_S.JPG
 1611134 9201_Overview_W01.JPG
 1797449 9202_Overview_W01.JPG
 1758585 9202_Overview_W02.JPG
 2844653 9203_Detail_TopTrench01.JPG
 2946303 9203_Detail_TopTrench02.JPG
 1499200 9203_Overview_W01.JPG
 1565892 9204_Overview_W01.JPG
 1840746 9205_Overview_S.JPG
 1773739 9206_Overview_W01.JPG
 1697405 9207_Overview_W.JPG
 1590857 9208_Overview_S.JPG
 1803886 9209_Overview_W.JPG
 1612263 9210_Overview_W01.JPG
 1927213 9211_Overview_W01.JPG
 1988413 9212_Overview_S01.JPG
 1918805 9213_Overview_S01.JPG
 1612445 9214_Detail_top.JPG
 1547729 9214_Overview_W01.JPG
 1580018 9215_Overview_E01.JPG
 1284625 9216_Overview_W01.JPG
 1410933 9217_Overview_W.JPG
 1967657 9218_Cremation.JPG
 1874539 9218_Cremation_DeerHorn.JPG
 1689437 9218_Cremation_Detail.JPG
 2135309 9218_Detail_Daub01.JPG
 1469091 9218_Detail_Daub02.JPG
 2082182 9218_Detail_ETrench.JPG
 1095026 9218_Overview_SE01.JPG
 2106227 9218_Pottery.JPG
 1412545 9219_Overview_W.JPG
 1255302 9220_Overview_S01.JPG
 1317988 9220_Overview_S02.JPG
 1337885 9222_Overview_Top.JPG
```

To find the 3 biggest files, I used the command `ls -S *.JPG` which lists all the files ending with ".JPG" by size from biggest to smallest. Now I can see that the three biggest files are "9254_Overview_SW.JPG", "9233_Overview_SE.JPG" and "9232_Overview_S.JPG".

```
[Emma-Maries-Air:hw5_photos Emma-Marie$ ls -S *.JPG
9254_Overview_SW.JPG                    9217_Overview_W.JPG
9233_Overview_SE.JPG                    9225_Overview_S02.JPG
9232_Overview_S.JPG                     9222_Overview_Top.JPG
9231_Overview_SE.JPG                    9220_Overview_S02.JPG
9253_Detail_Trench.JPG                  9216_Overview_W01.JPG
9253_Overview_S.JPG                     9220_Overview_S01.JPG
9229_Overview_W.JPG                     9225_Overview_S01.JPG
9233_Detail_Stratigraphy.JPG            9222_Overview_W02.JPG
9235_Overview_SE.JPG                    9223_Overview_S.JPG
9230_Detail_Trench.JPG                  9218_Overview_SE01.JPG
9233_Detail_Trench.JPG                  9222_Overview_W01.JPG
9233_Detail_Stratigraphy3.JPG           9226_Overview_E01.JPG
9260_Detail_Bunker.JPG                  9224_Overview_SW.JPG
9234_Overview_E.JPG                     9260_Detail_Drain.JPG
9233_Detail_Stratigraphy2.JPG           9260_Detail_Stratigraphy.JPG
9228_Overview_W3.JPG                    9260_Overview_E.JPG
9252_Overview_W.JPG                     9260_Overview_S.JPG
9228_Overview_W2.JPG                    9261_Overview_E.JPG
```

I can also use the command `ls -S *.JPG | head -n 3` to only list the three biggest files listed from the biggest to the smallest.

```
[Emma-Maries-Air:hw5_photos Emma-Marie$ ls -S *.JPG | head -n 3
9254_Overview_SW.JPG
9233_Overview_SE.JPG
9232_Overview_S.JPG
```

2) Some of the image files are empty, a sign of corruption. Can you *find* the empty photo files (0 kb size), count them, and generate a list of their filenames to make their later replacement easier?

To find the empty photo files I go to back one directory, so I am now in the directory containing the hw5_photos directory. Then I type the command `find hw5_photos -type f -empty` to find the empty files in the hw5_photos directory which are now listed in the

terminal.

```
[Emma-Maries-Air:Homework Emma-Marie$ find hw5_photos -type f -empty
hw5_photos/9289_Overview_S.JPG
hw5_photos/9318_overview_W.JPG
hw5_photos/9292_Overview_SE.JPG
hw5_photos/9306_ovrview_N.JPG
hw5_photos/9322_overview1_S.JPG
hw5_photos/9265_Overview_W.JPG
hw5_photos/9302_overview_W.JPG
hw5_photos/9278_Overview_W.JPG
hw5_photos/9287_Overview_S.JPG
hw5_photos/9322_RT(detail).JPG
hw5_photos/9317_overview_E.JPG
hw5_photos/9264_Overview_W.JPG
hw5_photos/9271_Overview_W.JPG
hw5_photos/9270_Overview_S.JPG
hw5_photos/9303_overview_W.JPG
hw5_photos/9319_overview_W.JPG
hw5_photos/9277_Overview_NW.JPG
```

I can count the files by adding using a pipe and then adding the wc function and the -l argument to the command: `find hw5_photos -type f -empty | wc -l` . The -l argument counts the number of lines which in this case is the same as the number of files. There are 73 empty files.

```
Emma-Maries-Air:Homework Emma-Marie$ find hw5_photos -type f -empty | wc -l
      73
```

To create a list of the empty files, I use the code `find hw5_photos -type f -empty > empty_files.txt` to create a text file containing the list of the empty documents. The arrow adds the findings to a new document which I chose to call “empty_files.txt”.

```
[Emma-Maries-Air:Homework Emma-Marie$ ls
HW2_monarchs.csv                Homework_3.pdf
HW2_monarchs.xlsx               Homework_4.Rmd
HW3_RStudio.R                   Homework_4.html
HW3_number2                     Homework_5.docx
HW3_task4.R                     SAFI_openrefine_hw2.csv
HW4_HomicideHistory             data_occupations_hw2.csv
Homework_1.docx                 empty_files.txt
Homework_1.pdf                  hw3_screenshot.png
Homework_2.docx                 hw5_photos
Homework_2.pdf                  kings.csv
Homework_3.docx                 ~$mework_5.docx
Emma-Maries-Air:Homework Emma-Marie$
```