5-W35 Managing Files on Steroids with Shell

Author: **Mina Almasi**
Date: 28/09-2022

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)

To identify the names and format of the three biggest files, the following commands are combined using pipes (|):

```
wc -c * | sort -nr | head -n 4| tail -n 3
```

The result is the following output:

```
(base) minaalmasi@Minas-MacBook-Pro-2 HW % wc -c * | sort -nr | head -
n 4| tail -n 3
 14761472 9240_Overview_S.RW2
 14733312 9247_Overview_SW.RW2
 14713856 9237_Overview_W.RW2
```

To gain a better understanding of the new command, it is broken into parts below:

1. The **wc** command is used along with the **-c** option to count the file size (amount of bytes). As we want the byte count for all files, a star * is written.

2. This is piped into the **sort** command with the **-nr** option to sort the list of files *numerically* (hence **-n**) but in *reverse* order (hence **-r**). By reverse order, descending is meant (i.e., from highest to lowest) as the default order when merely writing **-n** is ascending.

3. To get the three biggest files, we pipe into the command **head -n 4**. This would however also give us a "total size" line at the top of the list which is superflous.

4. To only display the three biggest files, we pipe into **tail -n 3**.

We notice that biggest file is 9240_Overview_S.RW2 with 14761472 bytes or ~14.8 megabytes. The format of all three files of the biggest files are .RW2.

(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?

The empty photo files can be found and counted with the following command:

```
find * -size 0 | wc -l
```

1. The **find** command is used with the **-size** option. It is specified that we only want to find files of size 0. This gives us a list of all files with a byte count of 0.

2. We then use the **wc** command with **-l** option to return the number of files in the list.

This gives the output:

```
(base) minaalmasi@Minas-MacBook-Pro-2 HW % find * -size 0 | wc -l
      73
```

Hence there are 73 files with a 0kb size.

To make their replacement easier later, we can save this list of files to a txt.file:

```
find * -size 0kb > empty_files.txt
```

To view the text file, we write.

```
cat empty_files.txt
```

This gives the output:

```
(base) minaalmasi@Minas-MacBook-Pro-2 HW % cat empty_files.txt
9260_Detail_Drain.JPG
9260_Detail_Stratigraphy.JPG
9260_Overview_E.JPG
9260_Overview_S.JPG
9261_Overview_E.JPG
9262_Overview_S.JPG
9263_Overview_W.JPG
9264_Overview_W.JPG
9265_Overview_W.JPG
9266_Overview_E.JPG
9267_Overview_W.JPG
9268_Overview_E.JPG
9269_Overview_S.JPG
9270_Overview_S.JPG
9271_Overview_W.JPG
9272_Overvieew_SE.JPG
9272_Overview_with_Dich_E.JPG
9273_Overview_S.JPG
9274_Overview_W.JPG
9275_Overview_SW.JPG
9276_Overview_SW.JPG
```

NB. Only a small snippet of the list is shown above.