Sample solution to HW 2: Other Utilities

Due: October 5

The second set of slides covers the `xargs` , `find` , `locate` , `grep` , `sed` , and `awk` commands.

1. Write a one-line command of piped Unix utilities (*only* the utilities from Lecture 2 and 3) to find all regular files on a system that have the same contents. Display the results in a list sorted from unique files to the largest group of files that are identical in content. All groups of identical files should be on the same line, comma separated. Explain how the command is able to accomplish this objective, step by step. Run this command on a small set of files with the following specification: 2 unique files, a group of 2 files with the same contents, a group of 4 files with the same contents.
   Sample command:
   Make sure there is no space after the line continuation character (\ )

   ```
   find "`pwd`" -type f | xargs md5sum | sort | tee ../md5out \
   | cut -f1 -d' '|sort|uniq -c | sort \
   | awk '{print $2}' |xargs -I{} bash -c "echo {}; grep -E -e '{}' ../md5out \
   | cut -d' ' -f3 | paste -s -d,; echo"
   ```

   Objective: List the files, hash the files, separate the checksum from the file name so you can count occurrences, search for those same checksums in the md5sum saved, removing the checksum from the filtered results, and finally transform the rows of the results into a comma separated list
   Commands Used:
   `xargs` , `find` , `grep` , and `awk` commands.
   Commands remaining:
   `locate` , `sed` commands.
   Input:

   ```
   --------------------------
   nigel@nigel-XPS-13-9343:~/cop4342/hw_2/tempfiles$ ls -1 *
   a
   b
   c
   d
   e
   f
   g
   nigel@nigel-XPS-13-9343:~/cop4342/hw_2/tempfiles$ cat *
   1
   2
   3
   3
   4
   ```

```
4
4
--------------------------
```

Output:

```
--------------------------
26ab0db90d72e28ad0ba1e22ee510510
/home/nigel/cop4342/hw_2/tempfiles/b

b026324c6904b2a9cb4b88d6d61c81d1
/home/nigel/cop4342/hw_2/tempfiles/a

6d7fce9fee471194aa8b5b6e47267f03
/home/nigel/cop4342/hw_2/tempfiles/c,/home/nigel/cop4342/hw_2/tempfiles/d

48a24b70a0b376535542b996af517398
/home/nigel/cop4342/hw_2/tempfiles/e,/home/nigel/cop4342/hw_2/tempfiles/f,/home/nigel/cop4342/hw_2/
--------------------------
```

2. Write a one-line command functionally equivalent to the problem above. Any commands from the list not used in problem 1 must be used here, but you may reuse commands.
   Functionally equivalent command:

```
locate "`pwd`" | xargs md5sum  | sort | tee ../md5out \
| cut -f1 -d' '|sort|uniq -c | sort | awk '{print $2}' \
|xargs -I{} bash -c "echo {}; sed -n -E -e '/{}/ p'  ../md5out \
| cut -d' ' -f3 | paste -s -d,; echo"
```

Objective: List the files, hash the files, separate the checksum from the file name so you can count occurrences, search for those same checksums in the md5sum saved, removing the checksum from the filtered results, and finally transform the rows of the results into a comma separated list
Commands Used:
`xargs` , `locate` , `sed` , and `awk` commands.
Output:

```
--------------------------
26ab0db90d72e28ad0ba1e22ee510510
/home/nigel/cop4342/hw_2/tempfiles/b

b026324c6904b2a9cb4b88d6d61c81d1
/home/nigel/cop4342/hw_2/tempfiles/a

6d7fce9fee471194aa8b5b6e47267f03
/home/nigel/cop4342/hw_2/tempfiles/c,/home/nigel/cop4342/hw_2/tempfiles/d

48a24b70a0b376535542b996af517398
/home/nigel/cop4342/hw_2/tempfiles/e,/home/nigel/cop4342/hw_2/tempfiles/f,/home/nigel/cop4342/hw_2/
--------------------------
```

3. Support the functional equivalence your answers above by comparing the outputs of the two commands with diff. Pipe each of the results to separate files, and show that the `diff` utility has no output.

Output:

```
nigel@nigel-XPS-13-9343:~/cop4342/hw_2/tempfiles$ diff ../command1_out ../command2_out
nigel@nigel-XPS-13-9343:~/cop4342/hw_2/tempfiles$
```