Ryan Clune
INFO 3401
Problem Set 2
2/11/2020

Problem Set 2

Part One: Shell Commands

1. Assume all of the directories below are valid directories (e.g., they exist and we have access to them). What would the following print out?

cd /usrs/INFO3401/../homework/./problem1

Pwd

Running this command (print working directory) would print the directory /usrs.homework/problem1

2. What set of commands would you use to achieve the following:

You've found yourself in a situation where your python program is trying to parse files in a directory and is throwing an error. You think it might be a permissions issue (i.e., you don't have permission to access the files in that directory) and need to verify the permissions on all files in that directory. The directory is in your home directory under the "datasets/activedata" directory. Make no assumptions about what directory you are currently working in.

Pwd

Cd ~

Cd datasets

Ls -1 activedata

Pwd

3. Briefly describe what the following set of commands would achieve. What process would happen and what would be printed to the command line?

cd ~ Brings you to your home directory

mkdir ./problem_set_1 Nothing is printed out. This will make a new directory called problem set 1.

touch submission.txt Nothing is printed out. This will make a text file called submission. This will be located in the home directory.

cd .. Nothing is printed out. This command takes you to the parent directory.

Pwd Print Working Directory

4. What set of commands would you use to achieve the following:

Copy a document called config.txt from your home directory to the root directory. Then, create a new directory in the root called preferences. Within preferences, copy prefs.txt from a matching directory called preferences in your home directory. Finally, determine that the start and end of the config.txt document in both root and home match.

Cp .~config.csv

Mkdir /preferences

Cp ~.preferences/prefs.txt

Head /config.txt

Head ~/config.txt

Tail /config.txt

Tail ~/config.txt

5. You have accidentally moved a file (~/Documents/datafile.csv) to your current directory rather than copying it. What command did you use to do that? What command should you have used to do that? What command might you use to undo it?

The person most likely used the mv (move) command. The command that should have been used would be cp (copy).

To fix this issue, I would use the cp command:

Cp ./datafile.csv ~/Documents/datafile.csv

6. Download and unzip asciify-master.zip from Canvas. Then, navigate to the directory and use the command line to run asciify.py. What does this script do? Include a screenshot of your results.

When I ran

Python3 asciify.py

I received the message:

Unable to find image in octocat.png

7. Install the Delorean Python package. What command did you use? Then, from the command line, launch Python. Use it to execute the following commands:

```
from delorean import Delorean

EST = "US/Eastern"

d = Delorean(timezone=EST)

print(d)
```

What prints to the command line?

To install this package, I used the pip command.

After executing the commands, Python returned the following:

Delorean(datetime=datetime.datetime(2020, 2, 11, 19, 24, 8, 850794), timezone='US/Eastern')

8. We can scrape webpage content or download collections of data files using wget. NASA, the USDA, and the NCBI all recommend using wget to download data from their repositories because it is faster and allows you to download data in bulk. We'll experiment with that on a smaller scale: use the command line to download data on the World Bank's budgetary expenditures at https://finances.worldbank.org/api/views/yu93-ayrw/rows.csv?accessType=DOWNLOAD

. What command did you use? Make sure to include this file as part of your submission.

Wget
https://finances.worldbank.org/api/views/yu93-ayrw/rows.csv?accessType=DOWNLOAD

9. The above budgetary data contains both individual items and aggregated totals across different sectors. Let's distill the data down to different aggregated totals. In this dataset, you can use the keyword "Total" to identify only those rows containing aggregate budget items. Use the command line to identify these rows. What command did you use? How many rows did this find (hint: the -c option can be helpful here)?

Grep Total rows.csv?accessType=DOWNLOAD.2

Grep Total rows.csv?accessType=DOWNLOAD.2 -c

Total = 33

(I accidentally downloaded the file 3 times in the process of trial and error, hence the ".2" after the download)

10. Now, print the lines containing the word "Total" to a file to "distilledExpenditures.csv". What command did you use? Include this file as part of your submission.

Grep Total rows.csv?accessType=DOWNLOAD.2 > distilledExpenditures.csv