Regex

Objectives

- Use Regex with common UNIX/Linux commands
- Practice using useful UNIX/Linux commands
    - diff
    - wc
    - cut
    - sed
    - awk
- Practice creating and running bash shell scripts
- Practice using pipes

Lab 2 Exercise

For each step please record the commands (and options) that you used to complete the task. At the end you will receive credit for the lab by showing your TA these commands.

Step 1: - **Download Practice Files from Moodle**

For Today's lab we will be using the following data files:

- fruitsOld.txt
- fruitsNew.txt
- testPasswd.txt
- grades.txt
- leetSpeak.txt
- regex_practice_data.txt

Download "Lab2_RequiredFiles.zip" from Moodle and copy the provided zip file to a directory on your machine (CSCI3308\Labs\) and decompress it using the unzip command:
unzip Lab2_RequiredFiles.zip -d Lab2
cd Lab2
Check to make sure each of the above files was correctly unzipped into the Lab2 directory.

Step 2: - **Use the diff command**

- Which "fruits" have been added to or removed from fruitsOld.txt to get to fruitsNew.txt.
- What do the '>' or '<' character mean at the beginning of each line in the output of diff command used?
- Try using the –c option with the above command, what does that do?

Step 3: - **Use the wc command**

- Find the # of lines in the testPasswd.txt
- Find the # of characters in the testPasswd.txt

Step 4: - **Use the cut command**

- Print a list of usernames from the testPasswd.txt file (print the first column only)
- Print out only the LN column and HW1 grade column from the grades.txt file

Step 5: - **Practice using pipes**

- Use cut, sort, and uniq to print out the groups that users are in within the testPasswd.txt file. (Each group is a #. When printing each group should get a different line and there should be no duplicates printed)
- Pipe the output of the above command into a file in your working directory.
- Use grep and cut to filter the testPasswd.txt file to only display usernames that start with 'm' or 'w' or 's' and their home directories. (sixth column)

Step 6: - **Use the sed command**

- Using sed and regular expressions try playing around with the leetSpeak.txt file.
    - Remove all the letters
    - Remove all the numbers
    - Replace all numbers with an '_'
- Create a **script** that pipes together multiple sed commands to replace each number with its matching character. Matching character is the character which closely resembles the number in shape. For example: 4 is a, 5 is s, 0 is o, 3 is e and 7 is t. (For this problem please use pipes) This can also be done without piping, how?
- It is possible that you may want to reuse this script on another file. How can you make it so that the script does not have to change each time you want to run it on a different filename?

Step 7: - **Use the awk command**

- Using the grades.txt file print out the first and last name of each student and calculate/print the grade in percentage that they currently have (assuming equal weights for each assignment).
- Using the grades.txt file calculate and print the class average for the lab assignment.

Step 8: - **More practice with regular expressions**

For the following problems use grep or egrep with the regex_practice_data.txt file.

1. How many phone numbers are in the dataset? (format ###-###-####)
2. How many city of Boulder phone numbers (e.g. starting with 303-441-…)?
3. How many email addresses? (an email address should contain @ character)
4. How many email addresses are from government domains (e.g. '.gov')?
5. How many email addresses are in 'first.last' name format AND involve someone who's first name starts with a letter in the first half of the alphabet?

Credit: To get credit for this lab exercise, submit your commands along with your final answers on Moodle as Lab2_<FirstName_LastName>.pdf.

For example, when a question asks: How many phone numbers, we want you to write the command along with the output which is the number of phone numbers.