

Homework Assignment 3
COP4342 Unix Tools (100 points+10 bonus points)
Due: 10/15/2019

1. (20 points + 10 bonus points) When we donate our email messages for research purpose, a key concern is privacy. For example, we would not like our personal information to be identified. A common process before donating email messages is to remove any sensitive information. In this problem, you need to develop a simple sed script to replace certain information in a given email folder (a file). The specific requirements are the following:
 - Replace duan@cs.fsu.edu by account@example.com
 - Replace cs.fsu.edu by example.com
 - Replace 128.186.120.196 by 192.100.100.100 (which is a private IP address)

An example run of the program is given below:

```
duan@linprog2 (~...homework/solutions) % spam.sed spam1 > spam1.output
duan@linprog2 (~...homework/solutions) %
```

Both files spam1 and spam1.output are provided to you.

- (10 bonus points) if your script can handle patterns over multiple lines (instead of only in individual lines). Two files, test1_sed and test1_sed.output, will be provided to you as sample input and output files. Note that your output file can be slightly different from the one provided here (you are allowed to format lines differently).
2. (80 points) A common task to write a book is to add an index page. In this problem, you need to write an awk (gawk) script to record the lines in which each word in the given file appears. The output should be of the following format: on each line, a word is first printed, followed by a colon “:”, then a space, and then the list of the line numbers where the word appears (separated by comma). If a word appears in a line multiple times, you should only report one time (for that line). An example run of an implementation is given below:

```
duan@linprog2 (~...homework/solutions) % index.awk test1.txt > test1.txt.output
duan@linprog2 (~...homework/solutions) %
```

Both files test1.txt and test1.txt.output will be provided to you.

- We ignore cases in generating the index page. For example, “Class” and “class” should be considered the same.
- Non-alphanumeric characters should not be included as part of a word. In particular, you should not include punctuation as part of a word. For example, if we have “test,” in the file, the word should be just “test”, instead of “test,”. You should read the awk reference manual to see how to handle this.

Submission: Tar your programs into a single file and submit online via Canvas (use the link following the assignment to submit). Make sure you tar your programs correctly. You are responsible for incorrect submissions (for example, empty tar file). You can untar the file **under a different directory** to make sure that you do include both programs in the tar file. Same late policy applies if the submission is incorrect and you need to submit a new version.

You should also be able to download your submission and verify if it is correct. If you cannot download the submission, contact the Canvas support team at FSU.