

CS 2204 Lab 4

your name here (please print):

your student ID number here:

WARM UP:

1. Create a subdirectory called 'lab4' under your home directory.
 2. Copy the file `soap.txt` into this directory. This file is either available from the directory `~cs2204/soap.txt` or at the URL: <http://courses.cs.vt.edu/~cs2204/fall2005/assignments/soap.txt>. If you are copying the file from the URL location, the **WRONG** way to save this file onto your `lab4` directory is to block and copy the text with your mouse (or the 'select all' option) and paste it into a file in, say `emacs`. The **RIGHT** way to save this file is to right click on the link and use the 'save link as' option. If you do it the first way, your copy-pasting might destroy the line formatting (e.g., one line could be broken down into multiple lines) and then your answers to the questions below will be incorrect.
 3. Before attempting the problems, make sure to set the environment variable `LC_ALL` to the value `C`. As explained in class, in the `bash` shell, you do this by: `export LC_ALL=C`. In the `tcsh` shell, you do this by: `setenv LC_ALL C`. Find out what shell you have and type the appropriate command.
-

QUESTIONS TO ANSWER:

Using `egrep`, find out the following information about `soap.txt`. For each of the questions below, write down your `egrep` command and also the answer.

1. (1 point) Does the string 'happy' appear in the file?
2. (1 point) How many lines contain at least one occurrence of the string 'soap'?
3. (3 points) How many lines contain multiple occurrences of the string 'soap'?

4. (2 points) How many lines do not begin with a lowercase letter?
5. (2 points) How many sentences end at the end of a line (i.e., the punctuation '.', '?', or '!' is the last character in the line)?
6. (3 points) How many lines contain a number? A number is a consecutive succession of one or more digits separated by 'white space' (e.g., a space, tab, newline etc.) characters on either side. So, the line 'My name is 234T' does not contain a number. But 'I have 3 oranges' does. Do a `man egrep` and you might find the bracket expression `[:space:]` useful.
7. (3 points) How many lines contain a two-digit number?
8. (2 points) How many lines contain a nine-letter word? Similar to the previous two problems, a word is a succession of (uppercase or lowercase or mixedcase) letters separated by whitespace on either side.
9. (3 points) What is the first character of the only line that contains a word with an even number of letters beginning with 'j'?