Filters and Regexps

1. What is your tutor's name, e-mail, how long have they been at UNSW, what are they studying, what is 1 interesting thing about them?

2. What are your class mates's names, what are they each studying, what is 1 interesting thing about each of them?

3. Are there marks for attending lectures, tutorials or labs?

4. What is an operating system? What operating systems are running in your tute room? What operating system do CSE lab computers run?

5.
   a. Write a regexp to match C preprocessor commands in a C program.
   b. Write a regexp to match all the lines in a C program except preprocessor commands
   c. Write a regexp to find line in a C program with trailing white space - one or white space at the end of line
   d. Write a regexp to match the names Barry, Harry, Larry and Parry
   e. Write a regexp to match a string containing the word `hello` followed later by the word `world`
   f. Write regexp to match the word `calendar` and all mis-spellings with 'a' replaced 'e' or vice-versa
   g. Write regexp to match a list of positive integers separated by commas, e.g. `2,4,8,16,32`
   h. Write regexp to match a C string whose last character is newline

6. Give five reasons why this attempt to search a file for HTML paragraph and break tags may fail.

   ```
   $ grep <p>|<br> /tmp/index.html
   ```

   Give egrep commands that will work.

7. For each of the regular expression below indicate how many different strings the pattern matches and give some example of the strings it matches. If possible these example should include the shortest string and the longest string.

   ```
   Perl
   ```

   ```
   Pe*r*l
   ```

   ```
   Full-stop.
   ```

   ```
   [1-9][0-9][0-9][0-9]
   ```

   ```
   I (love|hate) programming in (Perl|Python) and (Java|C)
   ```

8. This regular expression [0-9]*.[0-9]* is intended to match floating point numbers such as '42.5'. Is it appropriate?

9. What does the command `egrep -v .` print and why?
   Give an equivalent egrep command with no options, in other words without the -v and with a different pattern.

10. Write an egrep command which will print any lines in a file `ips.txt` containing an IP addresses in the range 129.94.172.1 to 129.94.172.25

11. For each of the scenarios below
    - describe the strings being matched using an English sentence
    - give a POSIX regular expression to match this class of strings

    In the examples, the expected matches are highlighted in bold.

    a. encrypted password fields (including the surrounding colons) in a Unix password file entry, e.g.

    ```
    root:ZHolHAHZw8As2:0:0:root:/root:/bin/bash
    jas:nJz3ru5a/44Ko:100:100:John Shepherd:/home/jas:/bin/bash
    ```

    b. positive real numbers at the start of a line (using normal fixed-point notation for reals, *not* the kind of scientific notation you find in programming languages), e.g.

    ```
    3.141 value of Pi
    90.57 maximum hits/sec
    half of the time, life is silly
    0.05% is the legal limit
    42 - the meaning of life
    this 1.333 is not at the start
    ```

    c. Names as represented in this file containing details of tute/lab enrolments:

```
    2134389|Wang, Duved Seo Ken         |fri15-spoons|
    2139656|Undirwaad, Giaffriy Jumis   |tue13-kazoo|
    2154877|Ng, Hinry                   |tue17-kazoo|
    2174328|Zhung, Yung                 |thu17-spoons|
    2234136|Hso, Men-Tsun               |tue09-harp|
    2254148|Khorme, Saneu               |tue09-harp|
    2329667|Mahsin, Zumel               |tue17-kazoo|
    2334348|Trun, Toyin Hong Recky      |mon11-leaf|
    2336212|Sopuvunechyunant, Sopuchue  |mon11-leaf|
    2344749|Chung, Wue Sun              |fri09-harp|
    ...
```

   d. Names as above, but without the trailing spaces (difficult). *Hint:* what are given names composed of, and how many of these things can there be?

12. Consider the following columnated (space-delimited) data file containing marks information for a single subject:

```
2111321 37 FL
2166258 67 CR
2168678 84 DN
2186565 77 DN
2190546 78 DN
2210109 50 PS
2223455 95 HD
2266365 55 PS
...
```

Assume that the student number occurs at the beginning of the line, that the file is sorted on student number, and that nobody scores 100.

   a. Give calls to the `sort` filter to display the data:
      i. in order on student number
      ii. in ascending order on mark
      iii. in descending order on mark
   b. Write calls to the `grep` filter to select details of:
      i. students who failed
      ii. students who scored above 90
      iii. students with even student numbers
   c. Write a pipeline to print:
      i. the details for the top 10 students (ordered by mark)
      ii. the details for the bottom 5 students (ordered by mark)
   d. Assuming that the command `cut -d' ' -f 3` can extract just the grades (`PS` , etc.), write a pipeline to show how many people achieved each grade (i.e. the grade distribution). E.g. for the above data:

```
    1 CR
    3 DN
    1 FL
    1 HD
    2 PS
```

13. Consider the following text file containing details of tute/lab enrolments:

```
    2134389|Wang, Duved Seo Ken         |fri15-spoons|
    2139656|Undirwaad, Giaffriy Jumis   |tue13-kazoo|
    2154877|Ng, Hinry                   |tue17-kazoo|
    2174328|Zhung, Yung                 |thu17-spoons|
    2234136|Hso, Men-Tsun               |tue09-harp|
    2254148|Khorme, Saneu               |tue09-harp|
    2329667|Mahsin, Zumel               |tue17-kazoo|
    2334348|Trun, Toyin Hong Recky      |mon11-leaf|
    2336212|Sopuvunechyunant, Sopuchue  |mon11-leaf|
    2344749|Chung, Wue Sun              |fri09-harp|
    ...
```

Assuming that the file is called `enrolments` , write pipelines to answer each of the following queries:

   a. Which tute is Hinry Ng enrolled in?
   b. How many different tutorials are there?
   c. What is the number of students in each tute?
   d. Are any students enrolled in multiple tutes?