CSC3320 System Level Programming
Homework 3
Due at 11:59 pm on Oct. 12th, 2014

Part1: Regular Expression (35 points in total)
Chose match(es) for each given regular expression (maybe multiple correct matches)
e.g. ‘ab+a’

a) ababa  b) aba  c)abba  d) aabbaa  e) aa

Answer: b, c (Because matched string should begin and end with ‘a’ and ‘b’ occurs at least once between leading and ending ‘a’)

[2 points for each in following 10 questions]

1) ‘a(ab)*a’
   (a) ababa  (b) aaba  (c) aabab  (d) aabbaa  (e) aa

2) ‘a(bc){1,5}’
   (a) abcbc  (b) abcb  (c) abb  (d) bcbc  (e) abcabc

3) ‘(a|b)*c’
   (a) abbc  (b) aabc  (c) abab  (d) c  (e) abb

4) ‘a.[bc]+’
   (a) azbc  (b) az  (c) azbcbc  (d) acc  (e) acz

5) ‘a.[0-9]’
   (a) azz  (b) a0z  (c) a01  (d) a0a  (e) aza

6) ‘[a-z]+[.\?!]’
   (a) good!  (b) Book.  (c) hard?  (d) cool? hot  (e) nothing

7) ‘[a-z]+[.\?!]s *[A-Z]’ (hint: s matches any space character and )
(a) Grade. A  (b) book. Z  (c) E. G  (d) Level. a  (e) index?

8) ‘(very )+(cool )?(good|bad) weather’
   (a) good weather  (b) very good weather  (c) cool weather
   (d) very cool bad weather  (e) cool good weather

9) ‘-?[0-9]+’
   (a) 3312  (b) -2231  (c) +32  (d) 0.5  (e) 2/3

10) ‘-?[0-9]*\.?[0-9]*’
   (a) 3312  (b) -2231  (c) +32  (d) 0.5  (e) 2/3

Write down the regular expression for following questions (3 points for each):

E.g. Social security number in the format of 999-99-9999.

Answer: [0-9]{3}-[0-9]{2}-[0-9]{4}

[3 points for each in following 5 questions]

1) Valid IP address for your computer (4 numbers separated by ‘.’).  
   (e.g. 323.645.0.1)

2) Valid URL beginning with “http://”. (e.g. http://cs.gsu.edu)

3) GSU panther ID in the format of 999-99-999. (e.g. 001-92-5434)

4) Valid email address, assuming ‘a-z’, ‘0-9’, ‘-’, ‘.’ are valid characters for user id and domain name has to be end with either “.com” or “.net”  
   (e.g. aabb-123.xy@g3.com, cdc0@12-3.net)

5) Phone number in any of the following format: 999-9999999,999-9999999,(999)-9999999.
Part 2: File Processing (20 points in total)
The following table is from Wikipedia. It shows the eleven highest mountains in Georgia.

<table>
<thead>
<tr>
<th>Mountain Name</th>
<th>Type of Location</th>
<th>Height in Feet</th>
<th>County</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brasstown Bald, (summit)</td>
<td>Union County</td>
<td>4784</td>
<td></td>
</tr>
<tr>
<td>Rabun Bald, (summit)</td>
<td>Rabun County</td>
<td>4696</td>
<td></td>
</tr>
<tr>
<td>Dick's Knob, (summit)</td>
<td>Rabun County</td>
<td>4620</td>
<td></td>
</tr>
<tr>
<td>Hightower Bald, (summit)</td>
<td>Towns County</td>
<td>4568</td>
<td></td>
</tr>
<tr>
<td>Wolfpen Ridge, (ridge high point)</td>
<td>Towns and Union Counties</td>
<td>4561</td>
<td></td>
</tr>
<tr>
<td>Blood Mountain, (summit)</td>
<td>Union County</td>
<td>4458</td>
<td></td>
</tr>
<tr>
<td>Tray Mountain, (summit)</td>
<td>Towns County</td>
<td>4430</td>
<td></td>
</tr>
<tr>
<td>Grassy Ridge, (ridge high point)</td>
<td>Rabun County</td>
<td>4420</td>
<td></td>
</tr>
<tr>
<td>Slaughter Mountain, (summit)</td>
<td>Union County</td>
<td>4338</td>
<td></td>
</tr>
<tr>
<td>Double Spring Knob, (summit)</td>
<td>Rabun County</td>
<td>4280</td>
<td></td>
</tr>
<tr>
<td>Coosa Bald, (summit)</td>
<td>Union County</td>
<td>4280</td>
<td></td>
</tr>
</tbody>
</table>

In above table, each line contains 5 fields separated by comma. Create a file named “mountainList.txt” in your computer, and copy the contents in above table.

Or use the following commands to get the file:

```
wget cs.gsu.edu/~ylong4/2014fall/csc3320/files/mountainList.txt
```

**Write down the command you would use for the following tasks.**

1) Use `grep` to output the mountains with ridge high point(2 points).

2) Use `egrep` to output the mountains which is a bald or knob(2 points).

3) Use `awk` to convert the value of height from feet to meter and provide a line number for each line(e.g. 2, Rabun Bald, (summit), 1431.34, meter, Rabun County). The result should be stored in ‘newMountainList.txt’. (3 points)
4) Based on ‘newMountainList.txt’, count the number of mountains belonging to Towns county. (hint: use **grep** and count the matched lines) (3 points)

5) Use **sed** to delete the second filed( including the ‘,’ after it) in each line of “mountainList.txt”. (hint: the second filed has a pair of parentheses and ‘[()]’ could match ‘(’).(3 points)

6) Use **awk** to count the number of mountains belonging to each county. And output the result to ‘countyMountain.txt’. The format of result should be:

   County name                      number of mountains
   (e.g. Union County  5)
   (hint: Assuming Wolfpen Ridge belongs to both of Town and Union County).(4 points)

7) **Sort** countyMountain.txt according to the number of mountains in descending order and output first two counties with more mountains.(3 points)