**CS 3421 Week 6 Problems: awk**

**1.** Given a comma-separated values (CSV) file, where fields are separated with commas and rows are separated with newlines, create an awk script which translates the contents to an html table. The entire table should begin with “<table>” and end with “</table>”. Each row should begin with “<tr>” and end with “</tr>”. Each cell within a row should begin with “<td>” and end with “</td>”. Also include the command to run the script called rec5-1.awk on a file named courses.csv (contents below) and print the output to courses.html.

**CSV file:**

Course Number,Section,Course Name,Start Time,End Time

CS3773,1,Software Engineering,10:00,10:50

CS3423,1,Systems Programming,11:00,11:50

CS3843,2,Computer Organization,2:00,2:50

# rec5-1.awk

BEGIN {

FS = ",";

print "<table>"

}

{

print "<tr>";

for(i=1; i<=NF; i++)

print "<td>"$i"</td>";

print "</tr>";

}

END {

print "</table>"

}

$ gawk -f rec5-1.awk courses.csv > courses.html

**2.** Create a one-line command using awk to replace the uniq utility. Executing the command on a given file should remove duplicate, consecutive lines. A sample input file is below.

# duplicated.txt

Hello world

Hello world

Hi

Guten tag

Hi

Hi

Hola

$ gawk 'prev != $0{print}; {prev = $0}' duplicated.txt

**3.** Create an awk script which takes a long list of files (i.e., the output of “ls -l”) and only prints lines belonging to the user passed in as a command line argument. At the end of the output, the script should print the average size, total size, and number of all zip files belonging to the user in the following format: “Zip file statistics for *USER:*” followed by the line: “Count: *XX* | Total size: *XX* | Average size: *XX*”. Keep in mind that the number of tokens per line may not be fixed if a file name contains spaces (which is bad practice anyway). Also include the command to run the script called rec5-3.awk and prints the output to stdout. An example list of files is included below.

$ls -l

-rw------- 1 rslavin faculty 120 Aug 28 2017 assignment1.zip

-rw------- 1 rslavin faculty 132450 Aug 28 13:29 bigfile.zip

-rw------- 1 clark faculty 12 Aug 28 13:29 file1.txt

-rw------- 1 rslavin faculty 540 Aug 28 13:29 file2.txt

-rw------- 1 clark faculty 708 Aug 28 13:29 file3.zip

-rw------- 1 rslavin faculty 70 Aug 28 13:30 input

drwxrwxr-x 2 rslavin faculty 70 Aug 28 13:30 more files

-rw------- 1 rslavin faculty 65 Aug 28 13:30 output

-rw------- 1 rslavin faculty 12 Aug 28 13:31 someScript.sh

-rw------- 1 rslavin faculty 673 Aug 28 13:29 test data.zip

# rec5-3.awk

BEGIN {

sum = 0;

count = 0;

}

{

if( match($3, arg1) ){

print $0;

if($NF ~ /zip$/){

count++;

sum+= $5;

}

}

}

END {

avg = 0

if( count > 0)

avg = sum / count;

print "Zip file statistics for ", arg1

print "Total: ", count, " | Total size: ", sum " | Average: ", avg

}

$ ls -l | gawk -f rec5-2.awk -v 'arg1=rslavin' -