**IS457**

**HW9**

**ClassID:39**

#!/bin/bash -x

# HW 9 - Due Monday Dec 3, 2018 in moodle.

# Upload .sh file to Moodle with filename: HW9\_457IDS\_YOURCLASSID.sh

# Please make sure all the commends work well in WholeTale, we will test your script.

# In your hard copy report, please include the UNIX / Linux script, input arguments, and results.

# There are multiple solutions for this homework. The grading will be based on the successful running of your code and the correct output as we specified. We will grade your homework on WholeTale.

# For this assignment we will use some basic commends of UNIX / Linux.

# The text Hw\_9.txt & adult.csv are uploaded to Moodle.

# You can use "text file" editor to edit HW9\_457IDS\_YOURCLASSID.sh and run in "Terminal".

# You can use the following commands to run the script (for example on google cloud):

# chmod +x HW9\_457IDS\_YOURCLASSID.sh

# ./HW9\_457IDS\_YOURCLASSID.sh Argument\_1 Argument\_2 Argument\_3 Argument\_4

# Here is a list of your input arguments:

# Argument\_1: a positive number

# Argument\_2: a lowercase word

# Argument\_3: text file ( .txt)

# Argument\_4: a positive integer which is less than 15

**In this case, the input arguments are: 11, command, Hw\_9.txt, 12**

# Q1 (2pts). Check whether your input integer(Argument\_1) is even or odd

# and print your result. (5 points)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q1 \*\*\*\*\*\*\*\*\*\*\*\*"

# Your answer here:

a=$(($1 % 2))

if [$a -eq 0]

then echo "$1 is even"

else echo "$1 is odd"

fi

\*\*\*\*\*\*\*\*\*\*\*\* Q1 \*\*\*\*\*\*\*\*\*\*\*\*

11 is odd

# Q2 (2pts). Input a lowercase letter(Argument\_2) and convert it to uppercase and print your result. (5 points)

# (Hint: tr)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q2 \*\*\*\*\*\*\*\*\*\*\*\*"

# Your answer here:

Upper=$(echo $2 | tr '[a-z]' '[A-Z]')

echo "Upper case of $2 is $Upper"

\*\*\*\*\*\*\*\*\*\*\*\* Q2 \*\*\*\*\*\*\*\*\*\*\*\*

Upper case of command is COMMAND

# Q3 (2pts). Convert the following phrase "CS 398/IS 457:INTRODUCTION

# TO DATA SCIENCE" into separate words, and put each word on its own

# line (ignoring space,'/' and ':'). (5 points)

# The output looks like:

# CS

# 398

# IS

# 457

# INTRODUCTION

# TO

# DATA

# SCIENCE

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q3 \*\*\*\*\*\*\*\*\*\*\*\*"

# Your answer here:

Var\_input="CS 398/IS 457:INTRODUCTION TO DATA SCIENCE"

Var\_output=$(echo $Var\_input | tr '[[:punct:][:space:]]' 'bbb')

Var\_output=${Var\_input//\ /\\n}

Var\_output=${Var\_output//:/\\n}

Var\_output=${Var\_output//\//\\n}

echo -e $Var\_output

\*\*\*\*\*\*\*\*\*\*\*\* Q3 \*\*\*\*\*\*\*\*\*\*\*\*

CS

398

IS

457

INTRODUCTION

TO

DATA

SCIENCE

# Q4 (2pts). Sort the answer in Q3 by descending order. (5 points)

# The output would look like:

# TO

# SCIENCE

# IS

# INTRODUCTION

# DATA

# CS

# 457

# 398

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q4 \*\*\*\*\*\*\*\*\*\*\*\*"

# Your answer here:

echo -e $Var\_output | sort -r

\*\*\*\*\*\*\*\*\*\*\*\* Q4 \*\*\*\*\*\*\*\*\*\*\*\*

TO

SCIENCE

IS

INTRODUCTION

DATA

CS

457

398

# Q5 (2pts). Count the lines of your text file(Argument\_3). (5 points)

# (Hint: wc)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q5 \*\*\*\*\*\*\*\*\*\*\*\*"

echo "The number of lines in $3 is:"

# Your answer here:

Var\_line=$(wc -l $3)

echo -e $Var\_line

\*\*\*\*\*\*\*\*\*\*\*\* Q5 \*\*\*\*\*\*\*\*\*\*\*\*

The number of lines in Hw\_9.txt is:

19 Hw\_9.txt

# Q6 (2pts). Count the frequency of a input word(Argument\_2) in a text

# file(Argument\_3), and print "The frequency of word \_\_\_\_ is \_\_\_\_ ".

# (5 points)

# (Hint: grep)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q6 \*\*\*\*\*\*\*\*\*\*\*\*"

echo "The frequency of word $2 is:"

# Your answer here:

Var\_fre=$(grep -o $2 $3 | wc -l)

echo $Var\_fre

\*\*\*\*\*\*\*\*\*\*\*\* Q6 \*\*\*\*\*\*\*\*\*\*\*\*

The frequency of word command is:

9

# Q7 (2pts). Print the number of unique words in the text file(Argument\_3).

# (5 points)

# (Hint: uniq, sort)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q7 \*\*\*\*\*\*\*\*\*\*\*\*"

echo "The number of unique words in the text file:"

# Your answer here:

grep -oE '\w+' $3 | sort | uniq -c | wc -l

\*\*\*\*\*\*\*\*\*\*\*\* Q7 \*\*\*\*\*\*\*\*\*\*\*\*

The number of unique words in the text file:

206

# Q8 (2pts). Print the number of words that begin with the letter 'b' in the

# text file(Argument\_3) (5 points).

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q8 \*\*\*\*\*\*\*\*\*\*\*\*"

echo "The number of words that begin with letter 'b':"

# Your answer here:

grep -oE '\bb(\w)\*' $3 | wc -l

\*\*\*\*\*\*\*\*\*\*\*\* Q8 \*\*\*\*\*\*\*\*\*\*\*\*

The number of words that begin with letter 'b':

7

# Q9 (2pts). Print top-k(Argument\_4) and find the most frequent word and their frequencies.

# (5 points).

# (Hint: head)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q9 \*\*\*\*\*\*\*\*\*\*\*\*"

echo "Top-$4 words are:"

# Your answer here:

grep -oE '\w+' $3 | sort | uniq -c | sort -nr | head -$4

\*\*\*\*\*\*\*\*\*\*\*\* Q9 \*\*\*\*\*\*\*\*\*\*\*\*

Top-12 words are:

48 the

20 and

14 shell

13 of

10 The

9 user

9 to

9 a

7 system

7 in

6 is

6 commands

# Q10 (4pts). The dataset adult-income.csv provides some clean records of adults to predict whether income exceeds $50K/yr. For details, visit the UCI repository.

# Calculate how many categories are there in "workclss" (2nd column) and print your result. (5 points)

# (Hint: awk)

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q10 \*\*\*\*\*\*\*\*\*\*\*\*"

Var\_result=$(awk -F, 'NR >= 2 {count[$2]++}END{for(i in count)print i":"count[i]"\\n"}' adult-income.csv)

echo -e $Var\_result

\*\*\*\*\*\*\*\*\*\*\*\* Q10 \*\*\*\*\*\*\*\*\*\*\*\*

Self-emp-not-inc:2541

Self-emp-inc:1116

Federal-gov:960

?:1836

State-gov:1297

Never-worked:7

Without-pay:14

Local-gov:2093

Private:22696

# Q11 (4pts). For your output in Q10, change the format of categories. Replace "-" with "\_".

# (Hint: sed)

# The output would look like:

# Private 22696

# Local\_gov 2093

# State\_gov 1298

# ...

echo "\*\*\*\*\*\*\*\*\*\*\*\* Q11 \*\*\*\*\*\*\*\*\*\*\*\*"

# Your answer here:

echo -e $Var\_result | sed -e 's/\-/\_/g'

\*\*\*\*\*\*\*\*\*\*\*\* Q11 \*\*\*\*\*\*\*\*\*\*\*\*

Self\_emp\_not\_inc:2541

Self\_emp\_inc:1116

Federal\_gov:960

?:1836

State\_gov:1297

Never\_worked:7

Without\_pay:14

Local\_gov:2093

Private:22696