



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

CS 251 Outlab3 - SED and AWK

Useful instructions for the Assignment

1. **Remember, you need to change the permission of the .sh file before executing it:**
`chmod u+x <script file> or chmod 777 <script file>`
2. **Follow the naming convention of the submission files strictly.**
3. **Remember to follow the exact output formats specified. Extra characters will lead to incorrect evaluation by the autograder.**
4. **Inside the script if you use any temporary file, write a command to remove it after the operation**



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

We have some assigned emailIDs belonging to an organization. We have a text file which contains email addresses of employees of an organization. You can mail any of them for your queries regarding organization. But recently we found out that some intruder has messed up this text file so now it contains invalid email addresses as well. So now if you want to email these people you need to know which are valid email addresses in this file. Use **sed** command to get the valid email addresses and store them in the **valid.txt** file.

Commands to run:

./q1.sh emails.txt

Your script should generate valid.txt file

emails.txt

alice98thomas.aitc@coap.edu

billi_brown@coap.edu

rsmith@bing.edu

james.charles@gmail.com

bella@will@yahoo.com

brown@mail@com

josh.charles@yahoo.com

bob@bing.com.edu

213040073.iitb.ac.in

213040073@iitb.ac.in

213040073@iitb.in

stella/james@coap.edu

josh..charles@bing.com

valid.txt

alice98thomas.aitc@coap.edu



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

~~stevan.james@csupres.edu~~
~~josh.charles@bing.com~~

Q2. a) [10 points]

'Hotel Orchid' at check in wants to store the entry of their guests. They have stored FirstName, LastName, and Timestamp at which they arrived in the hotel. Guest information is stored in a **guests.csv** file which has 3 columns namely FirstName, LastName, TimeStamp. Hotel wants to modify this guests.csv file such that it should add two columns namely Date and Time in it and then remove the TimeStamp column.

Format for timestamp column will be **<YYYY/MM/DD_HH:MM:SS>**. Date column should follow the **<DD MonthName YYYY>** format. Time column should show time in **12 hr** format.

For the entry, 2020/05/16_23:45:05 in Timestamp column you should write 16 May 2020 in Date column and 23:45:05 should be formatted as **11:45PM** in Time Column.

Your **final.csv** file should have 4 columns FirstName, LastName, Date, Time.

We have provided sample guests.csv and final.csv files please go through it.

Commands to run:

./q2a.sh guests.csv final.csv

Your script should take names of both input and output files as arguments.

Q2. b) [5 points]



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

e.g. img_12-31-06_06.jpg Rename this file name as 31-06-2012_06.jpg

Note:q2b.sh and images directory are in same folder

Q3. [10 points]

How about a simple challenge! Do you know about a tool called 'wc'? It helps you count the no. of lines, no. of words, characters and even bytes of a text file.

Can you design your own version of 'wc' without using 'wc' itself? Let us trivialize the problem a little. Write a bash script to take a filename as an input argument and calculate the no. of line, no. of words, and no. of characters. Your script should also be able to take flags which are optional and should be able to print simply the no. of line if the line flag is given as an argument. Note that only one flag will be provided at a time At minimum, the flags should be:

-lines: when this is provided to the script along with the input file, the script should ONLY print the no. of lines in the file.

-words: when this is provided to the script along with the input file, the script should ONLY print the no. of words in the file.

-chars: when this is provided to the script along with the input file, the script should ONLY print the no. of characters in the file (you must ignore the whitespaces) and count other characters.



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

```
=== begins here ===
```

```
    How are      you
    I am fine
    Thank you!
```

```
    How  are you?
```

```
=== ends here ===
```

Output:

```
>>./q3.sh <file>
```

```
>>35 characters, 11 words, 8 lines, 2 paragraphs
```

The above should be the output when nothing is provided as a flag to the script, but in case a flag is provided your script must only print the value for the flag provided.

For e.g.:

```
>>./q3.sh <file> -paras
```

```
>>2 paragraphs
```

Note: No typos in the output will be entertained.

For **sampleq3_input.txt** file, the output will be:

```
>>./q3.sh <file>
```

```
>>3611 characters, 630 words, 11 lines, 6 paragraphs
```

Q 4. [12 points]

Your task for this question is to perform addition of two numbers and write the result to file for each test case.



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

Input description:

First line of the file indicates the number of test cases. Second Line gives the base of the input and output separated by spaces. The next two lines of each test case have two numbers. The digits of the numbers are space separated. The base is guaranteed to be less than or equal 16.

Expected output for the above input:

1 0 0 0 1 0 0 1 0 1 1

Output description:

The digits in each result should be space separated. Each line should contain only one result.

Write the bash script q4.sh that takes two text files, one containing input and other to write the output. Make use of **awk** (mandatory!) to solve the task.

Q 5. [8 points]

AWK can also be used to treat separate files differently. In the following task you are supposed to write a command which takes in 3 parameters:

- <key>: A text file which has words separated by a hyphen. Note the first row is just the header row. It is ensured that there will not be two mappings of a single word.
- <replace>: A file which has words that need to be replaced with the words as given by the key. Note that



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

Note: Key will always be separated by the hyphen, the message itself will be present as words separated by <space>.

```
251-TA/sed-awk$ cat replace.txt
a aa b cc jkl mnopqrst
z qwed this is very secure encoding
good luck

251-TA/sed-awk$ ./Q5.sh key.txt replace.txt 20
you are genius how exactly did you do this question correct
e you didnt cheat (base) prathamesh@prathamesh-HP-Pavilion-x360-0

14-dh1xxx:~/Desktop/CS251-TA/sed-awk$ cat key.txt
_key_-_correctword_
a-you
aa-are
b-genius
cc-how
jkl-exactly
mnop-did
qrst-you
z-do
qwed-this
this-question
is-correctly
very-i
secure-hope
encoding-you
good-didnt
luck-cheat
```

Submission Instructions



CS251_Outlab3 - SED and AWK

Updated automatically every 5 minutes

```
tar -cvzf <roll_no>_outlab3.tar.gz <roll_no>_outlab3
```

The following is what your submission directory must look like. Please follow the submission format strictly.

```
<roll_no>_outlab3
├── q1.sh
├── q2a.sh
├── q2b.sh
├── q3.sh
├── q4.sh
├── q5.sh
└── references.txt
```

Note: references.txt should contain the exact links of webpages you referred to (apart from the ones linked in the resources shared before).

... May the Bash be with you! ...