An-Najah National University
Computer Engineering Department
Distributed Operation Systems - 10636456
Lab #0 – Part 2

REGEX/AWK

In this lab you will gain practical knowledge in working with Regular Expressions and some command-based data processing tools like AWK. You will be given a simple text file and asked to process the file to answer a number of required queries.

1- You need a linux-based environment to accomplish this task. You can either use any linux-based operating system (like Ubuntu for example), Cygwin, Windows Linux Subsystem or you can install docker and use the ubuntu container (see the recorded Containers Practical Lectures on moodle for a guide on how to do that).

2- Next skill you need is to learn is regular expressions (Regex). Regex is a way to describe a pattern (or filter) that describes a set of strings that matches a certain criteria or pattern. In other words, a regex accepts a certain set of strings and rejects the rest. Regex is widely used in programming, scripts, text processing, configuration, etc. It is a very powerful and useful skill to learn.

   To refresh/learn regex you have to finish the following interactive tutorial. You must finish all 16 lessons in the tutorial. I also encourage you to work on the practice problems as well. It takes around 15-30 minutes to finish this task.

   https://regexone.com/

3- Final skill we need to learn is AWK which is a command-based tool for text processing and data extraction and reporting that is also widely used for text data processing. To learn the basics of awk check out the following tutorial:

   https://linuxhandbook.com/awk-command-tutorial/

   This is a long tutorial so make sure to finish up to and including point 13 (you are encouraged to continue beyond that if you like).

   Pay special attention to point 8 in which you learn how to use AWK with Regular expressions.

   Once you feel confident about AWK try the following short AWK interactive tutorial:

   https://n8ta.com/projects/awk.html

   Also pay special attention to the part where you use regular expressions with AWK.

4- Now we are ready to start working on our main task. We will be processing the input file (data.txt) which you will find in the zip file associated with this homework. This file contains information regarding a number of bank accounts. For each account you have the first and last names of the owner, their city, his/her age and the amount of money in the account. Each line holds the information of one account.

   For each of the following requests (numbered from a to j) write down the awk command needed to get the correct answer.

a. Print only the first and last name of all the clients (there must be a space between first and last name).

The expected result looks like this:

```
firstname lastname
Herman Sanchez
Phil Parker
Bradie Garnett
Rudolf Crooks
Marcos Miller
Chad Garnett
Sally Evans
Chad Parker
```

b. Same as above but print the names are in reverse order (last name, first name) and make sure there is a comma between last name and first name.

```
lastname,firstname
Sanchez,Herman
Parker,Phil
Garnett,Bradie
Crooks,Rudolf
Miller,Marcos
Garnett,Chad
Evans,Sally
Parker,Chad
```

c. Print only first name and last name (without header that has the names of the columns: first name, last name, etc).

```
Herman Sanchez
Phil Parker
Bradie Garnett
Rudolf Crooks
Marcos Miller
Chad Garnett
Sally Evans
Chad Parker
```

d. Same as above but with numbers indicating the client order

```
1 Herman Sanchez
2 Phil Parker
3 Bradie Garnett
4 Rudolf Crooks
5 Marcos Miller
6 Chad Garnett
7 Sally Evans
8 Chad Parker
```

e. Print first and last names of customers who are more than 50 years old.

```
Herman Sanchez
Marcos Miller
Chad Parker
```

f. Print first and last name of customers who own more than 10000$

```
Chad Parker
```

g. Print the total sum of all money in all the accounts.

```
53900
```

h. Print all the information of all accounts whose owner first name is Chad. (hint. use REGEX)

```
Chad Garnett Miami   38 7420
Chad    Parker Seattle    55 12600
```

i. Print all the information of all accounts whose owner last name ends with the letter r (hint: REGEX)

```
Marcos Miller Seattle 66 4300
Chad Parker Seattle    55 12600
```

j. Print all the information of all accounts whose owner age has the number for the first and 2nd digits.

```
Rudolf  Crooks Miami 33 5800
Marcos Miller Seattle 66 4300
Chad Parker Seattle    55 12600
```