Programming Assignment 1

6 Possible Points

10/13/2023

Attempt 1

In Progress
NEXT UP: Submit Assignment

Add Comment

Unlimited Attempts Allowed

Details

Problem A - Get top K words (1 point)

Given a text file containing English words, return the top K most frequent words. You need to convert all the words to lowercase and ignore the stop words (provided in another text file).

- Your script should take 2 arguments:
  - the text file containing english words(Use sample.txt for testing)
  - k

Example:

sample.txt (Files -> Programming Assignment 1 -> sample.txt)

stopwords(Files -> Programming Assignment 1 -> stopwords)

```
Input :

File content: "The weather is sunny in SC. The weather is cloudy. the weather"

k:2
```

Expected output:

```
weather
cloudy
```

File to be submitted: *get_top_words.sh*

To test your code, use the following command: bash get_top_words.sh sample.txt k

Note: You can also run the above without using "bash", for example, ./get_top_words.sh

Note: If the top-k words have same frequency, then present them in alphabetical order.

Submit Assignment

Assume you have a web server or other application that adds a line to a log file each time it processes a request. Below are a few instances of lines from the log file (two lines are displayed below).

```
199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245

unicomp6.unicomp.net - - [01/Jul/1995:00:00:06 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
```

Here is the meaning of the fields of the above log lines. That is what is typically called metadata (or columns here). Note that "unknown" are columns that you can ignore for this assignment. And note that content_size is an int type, i.e., number of bytes. The columns of the log file are:

host, unknown1, unknown2, timestamp, method, url, version, response_code, content_size

- Your script should take 1 argument:
  - the log file (Use small_input.log for testing purpose)

Return the following:

- the average content size (treat "-" as 0 and round the final result) and
- the count of unique response codes.

Example:

```
Input:

small_input.log(Files -> Programming Assignment 1 -> small_input.log)
```

Expected output:

```
80554
11
```

File to be submitted: *log_analyzer.sh*

To test your code, use the following command: bash log_analyzer.sh small_input.log

Problem C - Get top k counties with highest percentage of fully vaccinated population (3 points)

For this part, you will need to download a data file. You can use the following command

```
wget https://raw.githubusercontent.com/UCSC-CSE-13S-02/asg1/main/covidVaccines.csv
```

Write a bash script to print the counties with the top-k highest percentage of fully vaccinated population, in decreasing order of fully vaccinated percentage. The `covidVaccines.csv` file has a column called "Completeness_pct" which has the information for vaccination percentages. Call the script **get_top_fully_vaccinated.sh** . Each line of output should contain the county name, the state, and fully vaccinated percentage, separated by a comma as shown below. If two counties have the same percentage, print them in alphabetical order

Submit Assignment

```
Santa Cruz County,AZ,94.9
Maverick County,TX,94.2
San Juan County,CO,94.2
```

```
Lares Municipio,PR,93.2
Irion County,TX,93.0
```

- Your script should take two command line arguments. The first is the .csv file and the second is k, the number of counties to output.
- Your script should print an error " Invalid number of arguments" if the number of arguments not 2.
- Your script should print an error "Cannot find/read covidVaccines.csv" if covidVaccines.csv is not an existing, readable file (your script should directly access covidVaccines.csv file).
- Your script should return county name, state and fully vaccinated percentage separated by comma.

File to be submitted: *get_top_fully_vaccinated.sh*

To test your code, use the following command: bash get_top_fully_vaccinated.sh covidVaccines.csv 7

What to submit?

Important - Please patiently read through and follow the instructions below and the instructions in the template file; otherwise, any violation of these instructions may result in 0 points for that problem.

This assignment has to be submitted on Canvas.

All the needed files are accessible at: Files -> Programming Assignment 1

For each problem submit one file as follows:

- Problem A : **get_top_words.sh**
- Problem B **: log_analyzer.sh**
- Problem C **: get_top_fully_vaccinated.sh**

Point Allocation (6 points):

- Problem A: 1 points for passing all the test cases
- Problem B: 2 points for passing all the test cases
- Problem C: 3 points for passing all the test cases
- For all problems, documentation (comments!) is necessary (you should always document your code, even if it's not required!)

View Rubric

Submit Assignment