

## Assignment 4 – Processing CSV Data with Threads

### Description:

This assignment encompasses three main objectives: implementing CSV processing functions to manage file operations, designing a data structure to analyze event type statistics, and integrating multithreading to concurrently process CSV data for performance evaluation with varying thread counts (1, 2, 4, and 8).

### Approach:

#### Step1) CSV Processing

1. Open the csv file through fopen in the csvopen function.
2. In the csvnext function, one line of string line is received in the linebuffer through getline.
3. Parse\_line\_buffer functions parse the corresponding line.
  - a. Check the characters one by one while turning the repetition statement by line\_buffer\_size.
  - b. If the current text is ""
    - i. Check double quotations and change double quotations to single quotations.
    - ii. If the following character == ',' and quotations\_mark\_cnt is an even number, it means the end of the string token (unless it is in the string). Therefore, it initializes quotations\_mark\_cnt to 0.
  - c. If the current character == ',' and quotations\_mark\_cnt is 0, (if the end of token)
    - i. tokenize
      1. If the current tokenization is a string, assign it to cur\_str except for "" at both ends.
        - a. If next\_line\_flag is true here, it means that there is no " at the front, so it is assigned to cur\_str except for the last ".
      2. If it's not a string, assign the entire token to cur\_str.
  - d. If the current character == '\n'
    - i. If quotations\_mark\_cnt is even (if it's the end of token)
      1. Tokenize in the same way as you did in c above.
    - ii. If not, it means that there is a '\n' in the middle of the string. Therefore, the tokens so far are put in prev\_str and a new line is received through getline. After that, parse the above process by repeating it in the same way.
  - e. If the current string token has '\n' in the middle
    - i. Prev\_str and cur\_str are combined and assigned to cur\_str.
4. Close the csv file through the csvclose function and free the memory.

## Issues and Resolutions:

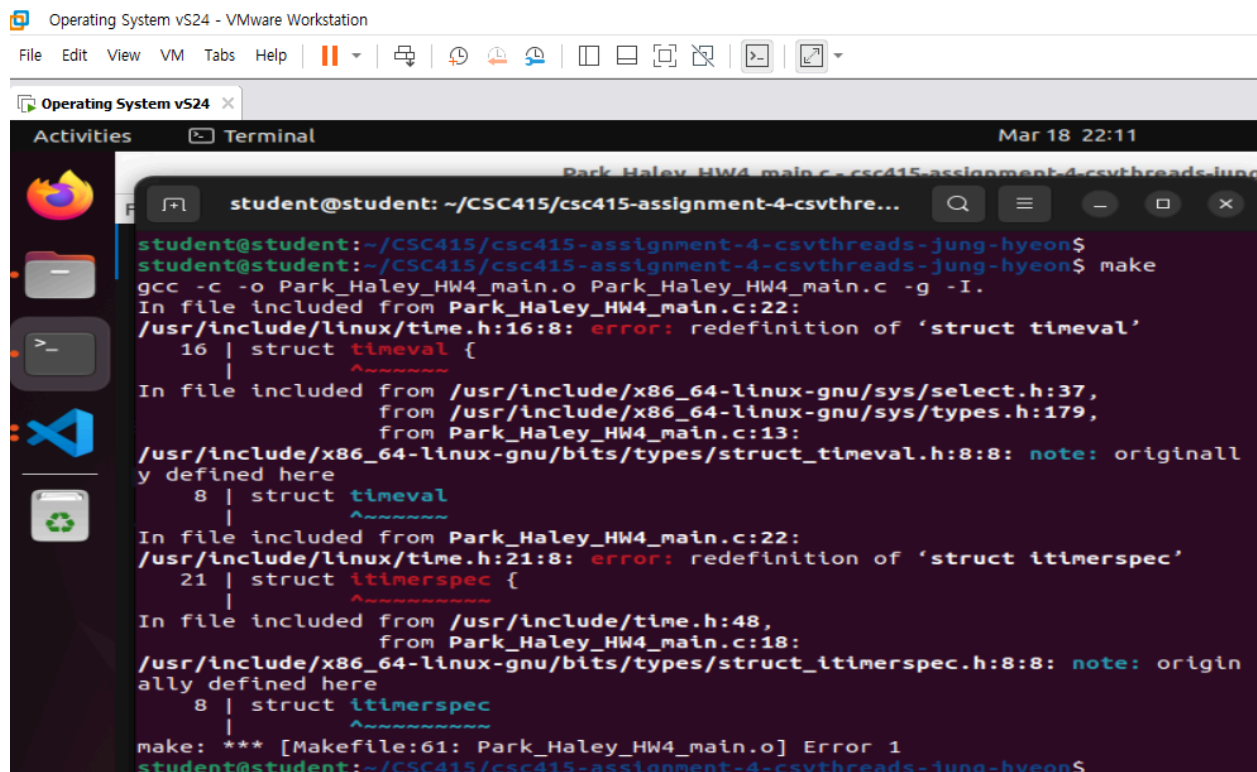
1. Garbage value

```
Data for line 2
Field 00: Smith, John
Field 01: 28dU
Field Incorrect - Should be: 28
Field 02: Green
Field 03: I want to have a multipl
Field Incorrect - Should be: I want to have a multiple
line comment
line comment"
```

At first, when allocating memory, I assigned it exactly to the length of the string. So the trash value appeared at the end. I solved it by putting the character '\0', which means the end of the string.

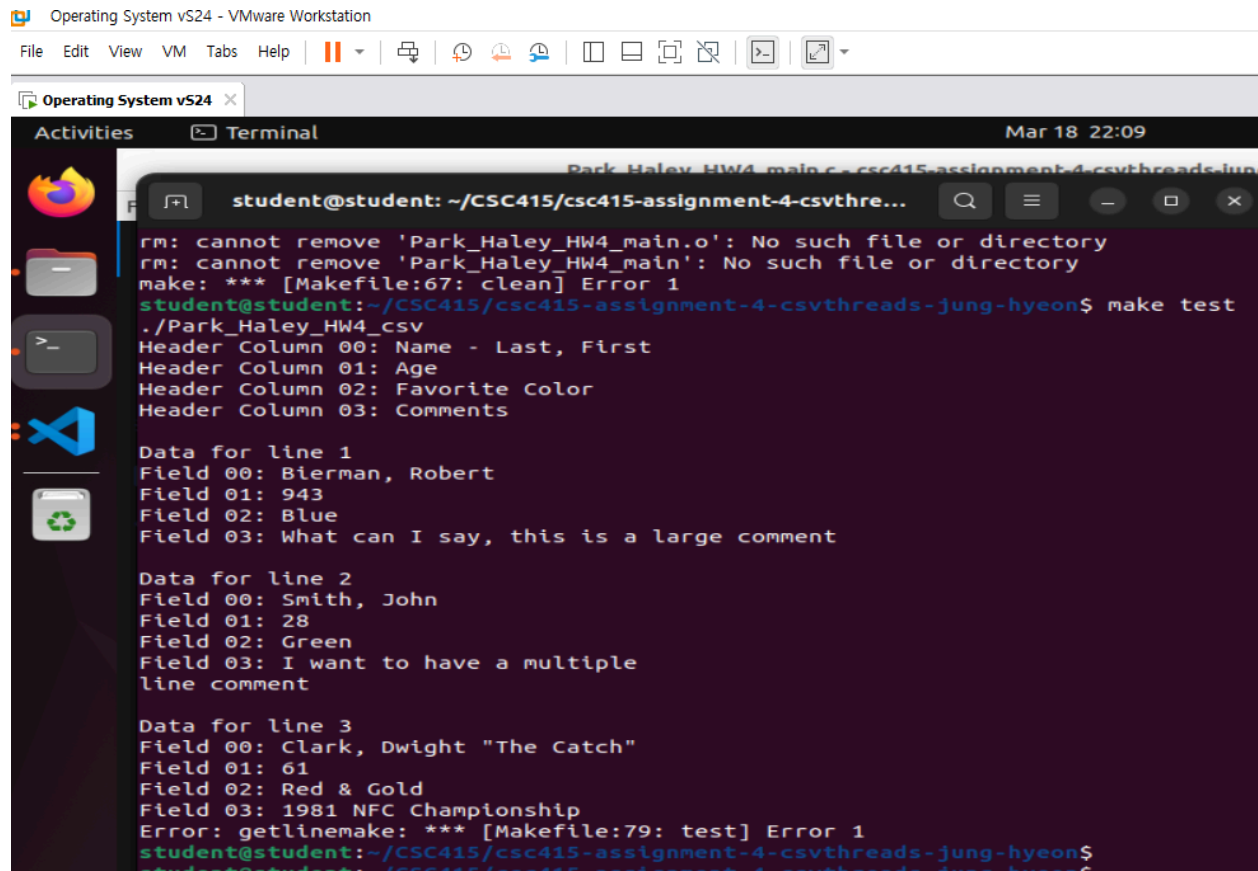
## Analysis:

## Screen shot of compilation:



```
Operating System vS24 - VMware Workstation
File Edit View VM Tabs Help
Operating System vS24 x
Activities Terminal Mar 18 22:11
student@student: ~/CSC415/csc415-assignment-4-csvthre...
student@student:~/CSC415/csc415-assignment-4-csvthre...$
student@student:~/CSC415/csc415-assignment-4-csvthre...$ make
gcc -c -o Park_Haley_HW4_main.o Park_Haley_HW4_main.c -g -I.
In file included from Park_Haley_HW4_main.c:22:
/usr/include/linux/time.h:16:8: error: redefinition of 'struct timeval'
16 | struct timeval {
    | ~~~~~
In file included from /usr/include/x86_64-linux-gnu/sys/select.h:37,
from /usr/include/x86_64-linux-gnu/sys/types.h:179,
from Park_Haley_HW4_main.c:13:
/usr/include/x86_64-linux-gnu/bits/types/struct_timeval.h:8:8: note: originall
y defined here
8 | struct timeval
  | ~~~~~
In file included from Park_Haley_HW4_main.c:22:
/usr/include/linux/time.h:21:8: error: redefinition of 'struct itimerspec'
21 | struct itimerspec {
    | ~~~~~
In file included from /usr/include/time.h:48,
from Park_Haley_HW4_main.c:18:
/usr/include/x86_64-linux-gnu/bits/types/struct_itimerspec.h:8:8: note: origin
ally defined here
8 | struct itimerspec
  | ~~~~~
make: *** [Makefile:61: Park_Haley_HW4_main.o] Error 1
student@student:~/CSC415/csc415-assignment-4-csvthre...$
```

## Screen shot(s) of the execution of the program:



```
Operating System vS24 - VMware Workstation
File Edit View VM Tabs Help
Operating System vS24 x
Activities Terminal Mar 18 22:09
student@student: ~/CSC415/csc415-assignment-4-csvthreads-jung-hyeon$ rm: cannot remove 'Park_Haley_HW4_main.o': No such file or directory
rm: cannot remove 'Park_Haley_HW4_main': No such file or directory
make: *** [Makefile:67: clean] Error 1
student@student:~/CSC415/csc415-assignment-4-csvthreads-jung-hyeon$ make test
./Park_Haley_HW4_csv
Header Column 00: Name - Last, First
Header Column 01: Age
Header Column 02: Favorite Color
Header Column 03: Comments

Data for line 1
Field 00: Bierman, Robert
Field 01: 943
Field 02: Blue
Field 03: What can I say, this is a large comment

Data for line 2
Field 00: Smith, John
Field 01: 28
Field 02: Green
Field 03: I want to have a multiple
line comment

Data for line 3
Field 00: Clark, Dwight "The Catch"
Field 01: 61
Field 02: Red & Gold
Field 03: 1981 NFC Championship
Error: getlinemake: *** [Makefile:79: test] Error 1
student@student:~/CSC415/csc415-assignment-4-csvthreads-jung-hyeon$
student@student:~/CSC415/csc415-assignment-4-csvthreads-jung-hyeon$
```