Assignment 2
Deadline: 11:59 PM on Nov 10

For Assignment 2, you will work in groups of up to three members. Ensure you form a different group from your assignment 1. If you prefer to work alone, that's acceptable as well.

Requirements
In this assignment, you are required to write a program that reads the content of a file and writes a sorted output to a file. The number of lines in the file is not known (it could be thousands of lines). The format of each line is:

```
>> cat input.txt
Mary Jackson Feb-2-1990 4.0 I 60
Jack He Feb-3-1990 2.45 D
Mike Johnson Sep-2-1980 3.125 D
Jane Zhang Mar-2-1970 3.8 I 120
```

You must use a merge sort to implement the sorting algorithm. Below are the sorting criteria where 1 is the first criteria to use and use the next criteria if the current criteria result in a tie.

1. Year of birth
2. Month of birth
3. Day of birth
4. Last name (alphabetical order)
5. First name (alphabetical order)
6. GPA
7. TOEFL (no TOEFL score is provided, then domestic students take precedence)
8. Domestic > International

The birthday has the month written in a string. Only the following abbreviation will be used for input and must be used for the output.

Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec

It must be in ascending order. For example, using the sorting criteria 1, someone born earlier must be outputted first. All sorting criteria must follow this ascending order. Below are additional specifications for this assignment.

1. An international student has an integer TOEFL score ranging from 0-120.
2. DomesticStudent struct does not have the TOEFL field.
3. I stands for an international student and D stands for a domestic student in the status column.

作业 2 截止日期：11 月 10 日晚上
11:59

对于作业 2，您将以最多三名成员的组进行工作。确保您与作业 1 形成不同的组。如果您更喜欢单独工作，那也是可以的。

要求
在这个作业中，您需要编写一个程序，读取文件的内容并将排序后的输出写入文件。文件中的行数未知（可能有数千行）。每行的格式是：

>> cat input.txt 玛丽·杰克逊 1990 年
2 月 2 日 4.0 我 60 杰克·赫 1990 年 2
月 3 日 2.45 D 迈克·约翰逊 1980 年 9
月 2 日 3.125 D 简·张 1970 年 3 月 2
日 3.8 我 120

您必须使用归并排序来实现排序算法。以下是排序标准，其中 1 是第一个使用的标准，如果当前标准导致平局，则使用下一个标准。

1. 出生年份
2. 出生月份
3. 出生日期
4. 姓氏（按字母顺序）
5. 名字（按字母顺序）
6. GPA
7. 托福（如果没有提供托福成绩，则优先考虑国内学生）
8. 国内 > 国际

生日的月份以字符串形式写出。输入和输出必须使用以下缩写。

一月, 二月, 三月, 四月, 五月, 六月, 七月, 八月, 九月, 十月, 十一月, 十二月

它必须按升序排列。例如，使用排序标准 1，出生较早的人必须优先输出。所有排序标准必须遵循这种升序。以下是此任务的其他规格。

1. 一名国际学生的托福成绩为 0-120 的整数。
2. DomesticStudent 结构体没有 TOEFL 字段。
3. 在状态栏中，I 代表国际学生，D 代表国内学生。

4. If the format of a student does not conform to the specified format, your program must write the appropriate message to an output file and exit ("Error: XXX"). replace XXX with an appropriate message.
5. The first argument is the input file, the second argument is the output file name, and the third argument is the option (specification 6). The run command will be ./<name of executable> <input file> <output file> <option>
6. The option field in the command line is an integer (1, 2 or 3). Here are descriptions of options.
   a. Option 1 only saves the sorted output of domestic students (no international students in the output file)
   b. Option 2 only saves the sorted output of international students (no domestic students in the output file)
   c. Option 3 only saves the sorted output of all students
7. You must handle corner cases including input format errors with the output message as in specification 4.
8. Assume that everyone has a first name and last name. No middle name. All names are case insensitive, but the output must have the same capitalization as the input.
9. GPA ranges from 0.0 to 4.3 and assume that the input only has up to 3 decimal places.
10. Birthday must be displayed with a three-letter abbreviation as shown above with the first letter being the capital.
11. Year of birth will range from 1950 to 2010

Restrictions

- For any reason, if your code does not compile/run, it will result in 0.
- If you use any standard library functions other than stdio.h, stdlib.h, string.You must consult with me or your dedicated lab instructor before using it.
- Every line must end with \n including the last line in the output with no extra trailing spaces.

Grading Criteria:

1. **Correctness (60 points)**
   - **(30 points)** Correctly implements merge sort according to the specified sorting criteria.
   - **(15 points)** Properly handles the differentiation between domestic and international students, including TOEFL scores.
   - **(15 points)** Generates accurate error messages for any improperly formatted input data
2. **Error Handling (10 points)**
   - **(5 points)** Handles all specified error cases correctly.
   - **(5 points)** Provides clear and accurate error messages.
3. **Code Quality (10 points)**
   - **(5 points)** Code is well-organized and follows best practices.
   - **(5 points)** Code is properly commented and easy to understand.

4. 如果学生的格式不符合指定格式，您的程序必须将适当的消息写入输出文件并退出（“错误：XXX”）。将 XXX 替换为适当的消息。

5. 第一个参数是输入文件，第二个参数是输出文件名，第三个参数是选项（规格 6）。运行命令将是 ./<可执行文件名> <输入文件> <输出文件> <选项>

6. 命令行中的选项字段是一个整数（1、2 或 3）。以下是选项的描述。 a. 选项 1 仅保存国内学生的排序输出（输出文件中没有国际学生） b. 选项 2 仅保存国际学生的排序输出（输出文件中没有国内学生） c. 选项 3 仅保存所有学生的排序输出

7. 您必须处理边缘情况，包括输入格式错误，并按照规范 4 中的输出消息进行处理。

8. 假设每个人都有一个名字和一个姓氏。没有中间名。所有名字不区分大小写，但输出必须与输入具有相同的大小写。

9. GPA 范围从 0.0 到 4.3，并假设输入仅有最多 3 位小数。

10. 生日必须以三字母缩写的形式显示，如上所示，首字母为大写。

11. 出生年份范围为 1950 年至 2010 年

限制

- 如果由于任何原因，您的代码无法编译/运行，将导致结果为 0。
- 如果您使用任何标准库函数，除了 stdio.h、stdlib.h 和 string 之外，您必须在使用之前咨询我或您的专门实验室指导老师。
- 每行必须以\n 结尾，包括输出中的最后一行，且没有多余的空格。

评分标准：

1. 正确性（60 分）
    - （30 分）根据指定的排序标准正确实现归并排序。
    - （15 分）妥善处理国内学生和国际学生之间的区别，包括托福成绩。
    - （15 分）为任何格式不正确的输入数据生成准确的错误消息
2. 错误处理 (10 分)
    - （5 分）正确处理所有指定的错误情况。
    - （5 分）提供清晰准确的错误信息。
3. 代码质量（10 分）
    - (5 分) 代码组织良好，遵循最佳实践。
    - （5 分）代码注释得当，易于理解。

4. **Adherence to Restrictions (10 points)**
    - **(5 points)** Does not use the strtok function.
    - **(5 points)** Only uses allowed standard library functions (stdio.h, stdlib.h, string.h) unless otherwise approved.
5. **Peer Evaluation (10 points)**
    - **(10 points)** Peer evaluation of group members' contributions and collaboration.

    Note: Not required if you are submitting individually.

Note: Any submission that does not compile or run due to not following instructions will receive a grade of 0 (no exceptions). Please refer to the rubric attached with the assignment in the Learning Hub for detailed guidelines and criteria.

Submission Files

- You must submit only one .c file named: a2.c (case sensitive).
- Take screenshots of the output and give them appropriate names.
- Submit the peer evaluation form along with the other files. If working alone, not required.
- Submission will be via Learning Hub.

4. Adherence to Restrictions (10 points)
    - (5 points) Does not use the strtok function.
    - (5 points) Only uses allowed standard library functions (stdio.h, stdlib.h, string.h) unless otherwise approved.
5. Peer Evaluation (10 points)
    - (10 points) Peer evaluation of group members' contributions and collaboration.

    Note: Not required if you are submitting individually.

Note: Any submission that does not compile or run due to not following instructions will receive a grade of 0 (no exceptions). Please refer to the rubric attached with the assignment in the Learning Hub for detailed guidelines and criteria.

Submission Files

- You must submit only one .c file named: a2.c (case sensitive).
- Take screenshots of the output and give them appropriate names.
- Submit the peer evaluation form along with the other files. If working alone, not required.
- Submission will be via Learning Hub.