

Bài tập lớn

Thiết kế và viết một chương trình phần mềm (Ngôn ngữ lập trình: C hoặc C++) sử dụng các hàm và cấu trúc dữ liệu phù hợp để mô phỏng một cảm biến nhiệt độ có thông số kỹ thuật sau:

- Dải đo: $-25^{\circ}\text{C} \div 125^{\circ}\text{C}$
- Độ phân giải: 0.2°C

Chương trình phần mềm thực hiện các nhiệm vụ sau:

1. Task 1: cho phép người dùng nhập chu kỳ trích mẫu và khoảng thời gian thực hiện phép đo (mô phỏng), chương trình cần tạo ra một tập hợp dữ liệu bao gồm các giá trị cảm biến một cách ngẫu nhiên. Nếu người dùng không nhập gì, chương trình sẽ sử dụng giá trị tần số trích mẫu và khoảng thời gian mặc định (tùy chọn, ví dụ: thực hiện mô phỏng phép đo trong 7 ngày, chu kỳ trích mẫu là 5 phút 1 mẫu). Tất cả các giá trị thu được phải lưu vào một file CSV (comma-separated values), chia thành 2 cột, một cột là thời gian (time) sử dụng định dạng 24 giờ và một cột chứa giá trị nhiệt độ (đơn vị: $^{\circ}\text{C}$), dòng đầu tiên là tiêu đề của mỗi cột. Số điểm đo tối thiểu = $1000 + 3$ số cuối của mã số sinh viên 1 trong 3 thành viên trong nhóm (tương ứng với số dòng, trừ dòng đầu tiên); có thể thử nghiệm với nhiều dữ liệu hơn: 5000, 10000, ... điểm đo. Ví dụ:

time	values
2021-04-16 08:30:00	25.5
2021-04-16 08:35:00	22.2
2021-04-16 08:40:00	26.1
2021-04-16 08:45:00	20.5

Tên file có định dạng như sau: “ho_va_ten_temperature_data.csv”, ví dụ: “nguyen_van_a_temperature_data.csv”.

Tham khảo thêm về file CSV: <https://www.ietf.org/rfc/rfc4180.txt>.

2. Task 2: sử dụng file CSV tạo ra ở task 1. Chương trình cần đọc file CSV nói trên và xử lý dữ liệu như sau:

- a. Task 2.1: giả sử cảm biến được đặt trong một môi trường có nhiệt độ thông thường chỉ nằm trong khoảng $15^{\circ}\text{C} \div 42^{\circ}\text{C}$, kiểm tra xem có dữ liệu cảm biến nào không hợp lệ không. Lưu các giá trị không hợp lệ này vào một file có tên “ho_va_ten_outlier_data.csv”, e.g.: “nguyen_van_a_outlier_data.csv”. Dòng đầu tiên là số lượng các điểm đo không hợp lệ; từ dòng thứ 2 trở đi là các giá trị đo có kèm theo thời gian với một dòng 2 là dòng tiêu đề tương tự như câu 1. Ví dụ file đầu ra:

invalid_data_points	
4	
time	values
2021-04-16 08:30:00	35.5
2021-04-16 08:35:00	15.1
2021-04-16 08:40:00	52.2
2021-04-16 08:45:00	10.2

- b. Task 2.2: Xác định giá trị nhiệt độ lớn nhất, nhỏ nhất và giá trị trung bình trong từng ngày (mô phỏng bao nhiêu ngày ở task 1 thì sẽ thống kê số ngày tương ứng). Lưu kết quả vào file csv: “ho_va_ten_data_summary.csv”, e.g.: “nguyen_van_a_data_summary.csv”

parameters	time	values
temperature_max	2021-04-16	38.2

temperature_min	2021-04-16	17.5
mean	2021-04-16	28.3
temperature_max	2021-04-17	41.7
temperature_min	2021-04-17	21.9
mean	2021-04-17	32.3
...		

c. Task 2.3: Đếm số lượng các giá trị nhiệt độ nằm trong các giải giá trị sau:

$15^{\circ}\text{C} \leq \text{temperature} < 18^{\circ}\text{C}$

$18^{\circ}\text{C} \leq \text{temperature} < 23^{\circ}\text{C}$

$23^{\circ}\text{C} \leq \text{temperature} < 28^{\circ}\text{C}$

$28^{\circ}\text{C} \leq \text{temperature} < 32^{\circ}\text{C}$

$32^{\circ}\text{C} \leq \text{temperature} < 35^{\circ}\text{C}$

$35^{\circ}\text{C} \leq \text{temperature} < 38^{\circ}\text{C}$

$38^{\circ}\text{C} \leq \text{temperature} \leq 42^{\circ}\text{C}$

Lưu kết quả vào một file có định dạng: “ho va ten_data_statistics.csv”:

range	quantity
15 to 18	4
18 to 23	6
23 to 28	11
28 to 32	16
32 to 35	8
35 to 38	5
38 to 42	1

d. Task 2.4: Sắp xếp dữ liệu ở file dữ liệu tạo ra trong Task 1 theo thứ tự giảm dần, lưu kết quả vào file “ho_va_ten_temperature_data_sorted.csv”, e.g.:

“nguyen_van_a_temperature_data_sorted.csv” trình bày như file

“nguyen_van_a_temperature_data.csv”. Có thể sử dụng bất kỳ thuật toán sắp xếp nào để

thực hiện nhưng phải tự viết chứ không được sử dụng các hàm sắp xếp có sẵn trong thư

viện. Hãy thử nghiệm 2-3 thuật toán sắp xếp khác nhau và so sánh thời gian thực hiện nhiệm

vụ sắp xếp của các thuật toán (có thể thử nghiệm với các số lượng dữ liệu khác nhau: 5000, 10000, ... điểm đo).

Lưu ý:

- Sinh viên làm theo nhóm tối đa 3 người.
- Các file dữ liệu, kết quả và file mã nguồn cần để chung trong một thư mục. Sinh viên cần xóa hết các file tạm ví dụ các file liên kết “*.o” và chỉ giữ lại các file quan trọng bao gồm mã nguồn, file cấu hình /makefiles và các file “*.csv”.
- Viết một báo cáo giải thích ý tưởng, thiết kế chương trình, lưu đồ thuật toán, cách thực hiện, kết quả và phân công nhiệm vụ ai làm phần nào. Báo cáo dài không quá 4 trang A4, viết theo mẫu của IEEE như file đính kèm.
- Khi nộp bài cần nén toàn bộ tất cả các file bao gồm cả file báo cáo gốc (file doc/docx nếu dùng words và các file latex nếu dùng latex). Tên file zip bao gồm mã số sinh viên và tên sinh viên của sinh viên đại diện và mã số sinh viên của các thành viên còn lại trong nhóm, ví dụ: “20170001_nguyen_van_a_20170002_20170003_miniproject.zip”
- Không copy source code / nội dung báo cáo từ các sinh viên khác / nhóm khác hoặc khóa khác. Nếu bị phát hiện copy bài sẽ bị coi như không nộp bài tập lớn.
- Nếu không nộp bài tập lớn sẽ bị trừ 2 điểm vào điểm quá trình. Nếu nộp muộn quá 1 ngày sẽ bị trừ 1 điểm, quá 1 tuần sẽ bị coi như không nộp.
- Thời hạn nộp bài: **27/05/2021 12h30 PM**
- Nộp bài trên Teams

- 3 bài nộp nhanh nhất hoàn thành đầy đủ nội dung với chất lượng tốt:
 - Thiết kế chương trình tốt, có tính sử dụng lại cao,
 - Trình bày chương trình sáng sủa, sạch đẹp
 - Có cách thức đặt tên hàm, biến, ... hợp lý, dễ hiểu, comments phù hợp
 - Chương trình chạy đúng và tin cậy
 - Báo cáo trình bày đúng theo mẫu, rõ ràng, dễ hiểu, đầy đủ nội dung.

sẽ được cộng 3 điểm vào điểm cuối kỳ.

- 4 bài nộp nhanh tiếp theo đáp ứng đầy đủ tiêu chí như trên sẽ được cộng 2 điểm
- 5 bài nộp nhanh tiếp theo nữa đáp ứng đầy đủ tiêu chí như trên sẽ được cộng 1 điểm
- Các bài được cộng điểm sẽ bảng trình bày thêm (nếu thời gian cho phép)
- Viết báo cáo bằng tiếng Anh (chất lượng chấp nhận được, không quá nhiều lỗi chính tả/ngữ pháp, và không dùng các chương trình tự dịch như google translate) và/hoặc sử dụng Latex được cộng thêm 1 điểm.