

- Tên học phần: **Lập trình Python**
- Mã học phần: **TIN1103** - Số tín chỉ: **3** - Đề số: **1**
- Thời gian làm bài: **120 phút** (Không kể thời gian phát đề)
- Loại đề: *Không được sử dụng tài liệu*

**Câu 1. (2 điểm) Chuẩn hóa dữ liệu**  
Chuẩn hóa dữ liệu là một bước quan trọng trong quá trình phân tích dữ liệu và học máy. Nó là quá trình chuyển đổi giá trị dữ liệu về chung một phạm vi, giúp giảm thiểu ảnh hưởng của biến lớn hơn so với biến nhỏ hơn. Một phương pháp phổ biến là chuẩn hóa min-max nhằm chuyển tất cả các giá trị dữ liệu về phạm vi mới  $[min, max]$ , với  $min < max$  là 2 giá trị cho trước. Phương pháp này thực hiện như sau. Cho vector  $X = [x_0, x_1, \dots, x_{n-1}]$  là biến dữ liệu cần chuẩn hóa,  $minX$  và  $maxX$  là giá trị cực tiểu và cực đại của vector  $X$ . Khi đó, vector chuẩn hóa  $new\_X = [new\_x_0, new\_x_1, \dots, new\_x_{n-1}]$  được tính như sau:

$$new\_x_i = \frac{x_i - minX}{maxX - minX}, \text{ với } i = 0, \dots, n-1.$$

Bạn hãy viết hàm *Chuanhoa()* để thực hiện công việc trên, với các yêu cầu:

- Hàm nhận 3 tham số đầu vào
  - $X$  là một *list* chứa vector dữ liệu,
  - $min$  và  $max$  là hai số thỏa mãn  $min < max$ , nhận giá trị mặc định  $min = 0$  và  $max = 1$ .
- Đầu ra của hàm
  - Vector  $new\_X$  được chuẩn hóa từ  $X$ .

Ví dụ:

- Gọi hàm: *Chuanhoa*([6, 3, 9, 8, 7], -1, 1)
- Kết quả trả về: [0.0, -1.0, 1.0, 0.33333333, 0.66666667]

**Câu 2. (2.5 điểm)** Để tính toán điểm môn Lập trình Python, chúng ta cần sử dụng hai từ điển để lưu trữ điểm QTHT và điểm thi:

- Từ điển thứ nhất: *key* là Mã sinh viên - *value* là điểm QTHT  
Ví dụ: {"23T1020054" : 8.5, "23T1080009" : 2.0, "23T1080012" : 7.9}
- Từ điển thứ hai: *key* là Mã sinh viên - *value* là điểm thi  
Ví dụ: {"23T1020054" : 6.5, "23T1080012" : 8.3}

Bạn hãy viết chương trình thực hiện các yêu cầu sau:



TRU  
PHC

- Tên họ
- Mã học
- Thời gian
- Loại đề

Câu 1 (1 điểm)  
phân chi  
255.255.2  
nhiều địa  
Câu 2 (1 điểm)  
CP  
Câu 3 (1 điểm)  
ng trong  
Câu 4 (1 điểm)  
Câu 5 (1 điểm)  
P-D sử  
u 6 (1 điểm)  
u 7 (1 điểm)  
u 8 (1 điểm)  
trong  
nh h  
9 (1 điểm)  
Dữ  
Đa  
10 (1 điểm)

- Nhập danh sách điểm QHTT và điểm thi của các sinh viên vào hai từ điển như mô tả ở trên.
- Tính tổng điểm của các sinh viên dựa vào từ điển thứ ba, biết  $\text{Tổng điểm} = \text{Điểm QHTT} \times 40\% + \text{Điểm thi} \times 60\%$ , các sinh viên không có điểm quá trình hoặc điểm thi sẽ có tổng điểm là 0. Sau đó, in ra tổng điểm của tất cả các sinh viên có mặt trong ít nhất một trong hai từ điển trên.

Câu 3. (3 điểm) Cho mô tả của hai lớp TacGia và Sach, như sau:

- Lớp TacGia gồm các thuộc tính: *Ten* (tên), *Email*, *GioiTinh* (giới tính)
- Lớp Sach gồm các thuộc tính: *TieuDe* (Tiêu đề), *DSTacGia* (danh sách tác giả), *NXB* (nhà xuất bản), *GiaBan* (giá bán).

Yêu cầu:

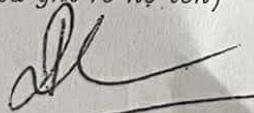
- Viết mã cho hai lớp *TacGia* và *Sach* như được mô tả, với các phương thức cơ bản gồm: `__init__()`, `__str__()`. Lưu ý, thuộc tính *DSTacGia* trong lớp *Sach* là một danh sách có kiểu *TacGia*.
- Nhập vào một danh sách các cuốn sách sau:
  1. Tiêu đề: "Python cơ bản"; tác giả: Bùi Việt Hà; NXB Đại Học Quốc Gia Hà Nội; Giá bán: 125.000vnd.
  2. Tiêu đề: "Python - Từ Cơ Bản Đến Giải Đề"; tác giả: Nguyễn Hoàng Trọng Lộc, Chung Long Hồ, Nguyễn Thanh Hùng; NXB Đại Học Sư Phạm Thành Phố Hồ Chí Minh; Giá bán: 92.000vnd.
  3. Tiêu đề: "Bài Tập Tin Học Dành Cho THCS"; tác giả: Phạm Thế Long, Bùi Việt Hà, Bùi Văn Thanh, Ngô Ánh Tuyết, Nguyễn Chí Trung; NXB Giáo dục Việt Nam; Giá bán: 21.000vnd.
- Nhập vào tên của một tác giả và in ra tất cả các cuốn sách mà người đó có trong danh sách tác giả.

Câu 4. (2.5 điểm) Để chuẩn bị cho việc xây dựng một ứng dụng xử lý ngôn ngữ tự nhiên, Bi cần một chương trình tiền xử lý các file văn bản tiếng Việt. Bạn hãy giúp Bi viết một chương trình thực hiện lần lượt các công việc cụ thể sau đây:

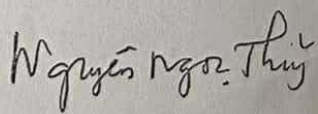
- Đọc file dữ liệu *data.txt* (tải xuống từ Google Classroom)
- Tách từng câu trong văn bản vừa đọc đưa vào một danh sách.
- Xóa tất cả các ký tự đặc biệt trong các câu, ở đây ký tự đặc biệt được hiểu là những ký tự không phải là chữ cái, số hoặc khoảng trắng.
- Chuyển tất cả các ký tự in HOA thành in thường.
- Ghi tất cả các dòng đó ra 1 file mới *preprocessed\_data.txt*.

(Ghi chú: Cán bộ coi thi không được giải thích gì thêm)

DUYỆT  
(Ký và ghi rõ họ tên)

  
Nguyễn Đại Bình

ĐẠI DIỆN NHÓM CÁN BỘ RA ĐỀ  
(Ký và ghi rõ họ tên)

  
Nguyễn Ngọc Thùy