

KIỂM TRA GIỮA KỲ

Môn Lập trình Python

Thời gian 90 phút

Đề 2

Bài 1 (5 điểm)

Biểu diễn vector thưa. Trong thực tế có nhiều bài toán làm việc với dữ liệu là các vector với kích thước rất lớn, nhưng đa phần các thành phần trên vector này đều có giá trị bằng 0, chỉ có một số ít các thành phần có giá trị khác không. Do vậy, thay vì biểu diễn vector đầy đủ với rất nhiều vị trí bằng 0, $\mathbf{t} = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 5, 0, 0, 0, 9, 0]$ ta có thể biểu diễn vector bằng cách chỉ biểu diễn thông tin kích thước vector và các thành phần khác không của vector. Như ví dụ trên, ta có thể biểu diễn \mathbf{t} bằng 1 bộ (tuple) 2 phần tử $\mathbf{spr} = (16, \{5:1, 10:5, 14:9\})$ biểu diễn kích thước vector $\mathbf{t} = 16$, và các thành phần khác 0 (đánh chỉ số từ 0): vị trí 5 = 1, vị trí 10 = 5, vị trí 14 = 9.

Hoàn thiện các hàm trong file **SPR.py** theo các yêu cầu sau:

1. Hàm **sparseForm(t)**: Hàm này tính và trả lại biểu diễn (spr) thưa của vector \mathbf{t} theo mô tả trên.
2. Hàm **revert(spr)**: Hàm này trả về vector tương ứng với biểu diễn thưa spr.
3. Hàm **dot(spr1, spr2)**: Thực hiện tính và trả lại tích vô hướng của 2 vector \mathbf{u}, \mathbf{v} được biểu diễn bởi biểu diễn thưa spr1 và spr2.

Trong đó tích vô hướng $(\vec{u} \cdot \vec{v}) = \mathbf{u}_1\mathbf{v}_1 + \mathbf{u}_2\mathbf{v}_2$, trong đó, $\mathbf{u} = (u_1, u_2)$, $\mathbf{v} = (v_1, v_2)$. Nếu véc-tơ có nhiều hơn hai thành phần, bạn chỉ việc thêm tiếp $+ u_3v_3 + u_4v_4 \dots$

4. Hàm **getCosineSim(spr1, spr2)**: thực hiện tính và trả lại giá trị cosine của góc giữa 2 vector \mathbf{u}, \mathbf{v} được biểu diễn tương ứng bởi 2 biểu diễn thưa spr1 và spr2.

Cosine của góc giữa hai vector được tính theo công thức sau:

$$\cos(\vec{u}, \vec{v}) = \frac{(\vec{u} \cdot \vec{v})}{(\|\vec{u}\| \times \|\vec{v}\|)}$$
$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2}$$

Bài 2 (5 điểm)

Hoàn thiện các phương thức trong file **TFIDF2.py** theo các mô tả trong file.

-----Hết-----

Chú ý: Sinh viên đọc cẩn thận các lưu ý sau đây để tránh mắc các lỗi khi làm bài

1. Sinh viên chỉ cần hoàn thiện các phương thức được yêu cầu, tuyệt đối không chỉnh sửa nguyên mẫu (bao gồm danh sách đối và kiểu dữ liệu trả về) của các phương thức này, chỉ viết thêm vào phần nội dung phương thức.
2. Sinh viên có thể viết thêm các phương thức phụ trợ khác nếu cần, tuy nhiên tốt nhất tránh đặt tên trùng với tên các phương thức đã cho.
3. Một số phương thức được chấm điểm độc lập, đúng phương thức nào sẽ được điểm phương thức đó, tuy nhiên cần đảm bảo bài làm có thể biên dịch và chạy được. Các bài có lỗi biên dịch, hoặc không chạy được sẽ nhận điểm 0.
4. Các phương thức sinh viên chưa làm đến, hoặc không hoàn thiện được thì để lại như nguyên mẫu ban đầu, không chỉnh sửa, thêm bớt, tránh việc bài không biên dịch được.
5. Khi làm bài sinh viên có thể tự nộp bài và chấm điểm trên 1 bộ test mẫu, kết quả này chỉ là kết quả tham khảo, bài làm của sinh viên sẽ được chấm lại trên một bộ test khác và thông báo kết quả sau.