

Cắt ghép xâu (100 điểm)

Tin sinh học (Bioinformatics) là một lĩnh vực khoa học sử dụng các công nghệ của các ngành để giải quyết các vấn đề sinh học. Một bài toán được nghiên cứu để sử dụng phân tích dữ liệu sinh học như sau: Cho hai xâu S_1, S_2 , hãy tìm cách cắt xâu S_1 thành ít đoạn nhất sau đó ghép tất cả các đoạn lại theo một thứ tự nào đó để nhận được xâu S_2 .

Ví dụ: $S_1 = "abcabc", S_2 = "aabbcc"$, một phương án cắt S_1 tại các vị trí sau kí tự thứ 1, thứ 3, thứ 5 nhận được bốn xâu "a", "bc", "ab", "c" để ghép được xâu $S_2 = "aabbcc"$.

Dữ liệu: Vào từ thiết bị vào chuẩn có khuôn dạng:

- Dòng đầu chứa xâu S_1 chỉ gồm các kí tự 'a' đến 'z';
- Dòng thứ hai chứa xâu S_2 chỉ gồm các kí tự 'a' đến 'z'.

Kết quả: Ghi ra thiết bị ra chuẩn theo khuôn dạng: Ghi -1 nếu không tồn tại cách cắt thỏa mãn; ngược lại, gọi c là số điểm cắt và các vị trí cắt là $1 \leq p_1 < p_2 < \dots < p_c < n$, đánh số các đoạn được cắt ra theo thứ tự từ đầu đến cuối bắt đầu từ 1 đến $c + 1$, trường hợp này cần ghi ra ba dòng theo khuôn dạng:

- o Dòng đầu chứa số nguyên dương c ;
- o Dòng thứ hai gồm c số nguyên là các vị trí cắt p_1, p_2, \dots, p_c ;
- o Dòng thứ ba gồm $c + 1$ là một hoán vị của $1, 2, \dots, c + 1$ mô tả cách ghép các đoạn, các số cách nhau bởi dấu cách.

Ràng buộc:

- Có 30% số test ứng với 30% số điểm có độ dài xâu S_1, S_2 đều không vượt quá 20;
- Có 30% số test khác ứng với 30% số điểm có độ dài xâu S_1, S_2 đều không vượt quá 60;
- Có 40% số test còn lại ứng với 40% số điểm có độ dài xâu S_1, S_2 đều không vượt quá 10^4 .

Cách tính điểm:

Có 20 test, mỗi test 5.0 điểm. Gọi số điểm cắt do thí sinh tìm được là c , số điểm cắt của Ban giám khảo là q , khi đó số điểm bạn đạt được cho mỗi test là $5.0 \times \min\{1, \frac{q}{c}\}$.

Ví dụ:

Dữ liệu vào	Kết quả ra
abcabc	3
aabbcc	1 3 5
	1 3 2 4