



DNA Prefix

Link submit: http://lightoj.com/volume_showproblem.php?problem=1224

Solution:

C++	https://ideone.com/R1eE7Z
Java	https://ideone.com/eMBQDj
Python	https://ideone.com/kB4R9A

Tóm tắt đề:

Các chuỗi DNA gồm các ký tự A, C, G, T ghép lại với nhau. Cho bạn n chuỗi DNA tìm những chuỗi DNA nào có tiền tố giống nhau. Sau đó, bạn sẽ làm phép tính “số ký tự tiền tố giống nhau” nhân “số lượng chuỗi có tiền tố giống nhau”. Trong n chuỗi sẽ có nhiều loại như vậy, bạn hãy tìm loại sao cho khi làm phép tính trên ra kết quả lớn nhất.

Input:

Dòng đầu tiên chứa số lượng bộ test T ($T \leq 10$), mỗi bộ test gồm có:

- Số lượng chuỗi n ($1 \leq n \leq 50.000$)
- n dòng tiếp theo mỗi dòng chứa một chuỗi.

Output:

In ra “Case x: ans” với x là số thứ tự của bộ test. ans là kết quả lớn nhất nhận được khi thực hiện phép tính.

Ví dụ:

3 4 ACGT ACGTGCGT ACCGTGC ACGCCGT 3 CGCGCGCGCGCGCCCCGCCCCGCGC CGCGCGCGCGCGCCCCGCCCCGCAC CGCGCGCGCGCGCCCCGCCCCGCTC 2 CGCGCCGCGCGCGCGCGCGC GGCGCCGCGCGCGCGCGCTC	Case 1: 9 Case 2: 66 Case 3: 20
---	---------------------------------------

Giải thích ví dụ:

Ví dụ gồm 3 bộ test. Trong bộ đầu tiên có 4 chuỗi.

Nếu chọn tiền tố là "AC" thì có 2 chuỗi đều có vậy $\text{ans} = 2 \times 4 = 8$.

Nếu chọn tiền tố là "ACG" thì có 3 tiền tố $\text{ans} = 3 \times 3 = 9$.

Kết quả in ra 9.

Hướng dẫn giải:

Bài này dùng Trie, mỗi nút lưu thêm một biến dùng để đếm số lượng nút lá của nút ở tầng hiện tại. Lấy biến này nhân với tầng hiện tại là sẽ ra kết quả.

Độ phức tạp: $O(T * \text{string_length} * N)$ với T là số lượng bộ dữ liệu, string_length là độ dài tối đa của mỗi chuỗi ADN trong dataset và N là số lượng chuỗi ADN.