

HASHING STRING

A. LÝ THUYẾT

1. Phép băm (Hashing)

Cho xâu ký tự độ dài n : $S = s_1s_2 \dots s_n$ trong đó s_i là ký tự mã ASCII (có giá trị trong khoảng 0...255).

Chọn một số nguyên dương $B > 255$ ta có thể xem xâu S như là một số nguyên biểu diễn trong cơ số B :

$$h(S) = s_1 \cdot B^{n-1} + s_2 \cdot B^{n-2} + \dots + s_{n-1} \cdot B^1 + s_n \cdot B^0$$

Và các phép toán xử lý xâu có thể đưa về các phép toán xử lý số.

Tuy nhiên dễ nhận thấy $h(S)$ có giá trị rất lớn nên trong thực tế việc sử dụng trực tiếp giá trị $h(S)$ thường là không thể hoặc nếu có thể thì thường không hiệu quả (về mặt thời gian).

Một trong những giải pháp khắc phục hạn chế trên là thay vì dùng trực tiếp giá trị $h(S)$ ta chọn một số nguyên dương P sau đó đặt:

$$h(S) = (s_1 \cdot B^{n-1} + s_2 \cdot B^{n-2} + \dots + s_{n-1} \cdot B^1 + s_n \cdot B^0) \% P$$

Giá trị $h(S)$ nêu trên còn được gọi là giá trị băm của xâu. Tuy vậy có thể thấy sẽ phát sinh tình huống "va chạm" tức là có hai xâu khác nhau nhưng có cùng giá trị băm trùng nhau. Để hạn chế điều thông thường sẽ chọn:

- B là số nguyên tố (ví dụ $B = 331$)
- P là số nguyên tố đủ lớn (ví dụ $P = 1000000009$)

Ngoài ra thay vì sử dụng một số nguyên để băm ta có thể sử dụng một cặp hai số nguyên. Số thứ nhất là phần dư khi chia cho số nguyên tố P_1 còn số thứ hai là phần dư khi chia cho số nguyên tố P_2

2. Tính mảng băm xâu:

Cho xâu $S = s_1s_2 \dots s_n$ hãy lập mảng $H[1], H[2], \dots, H[n]$ trong đó $H[i]$ là giá trị băm của xâu $s_1s_2 \dots s_i$

```
H[0]=0;
for(int i=1;i<=n;++i) {
    H[i]=(1LL*H[i-1]*B+s[i]) % P;
}
```

3. Tính giá trị băm của dãy con $s_i s_{i+1} \dots s_j$

Đặt $LT[i]=B^i \% P$. Ta có thể tính trước mảng này như sau:

```
LT[0]=1;
for(int i=1;i<=n;++i) LT[i]=(1LL*LT[i-1]*B) % P;
```

Khi đó hàm lấy giá trị băm của xâu con $s_i s_{i+1} \dots s_j$ có thể viết

```
int Hash(int i,int j) {
    int res=((H[j]-1LL*H[i-1]*LT[j-i+1]) % P + P) % P;
    return res;
}
```

4. Phép băm tập hợp

Cho $S = \{x_1, x_2, \dots, x_n\}$ là tập hợp với $0 < x_i < n$ khi đó ta có thể tương ứng tập hợp này với số nguyên:

$$h(S) = (B^{x_1} + B^{x_2} + \dots + B^{x_n}) \% P$$

Trong đó B là một số nguyên tố ($B > n$) và P cũng là một số nguyên tố (giống như trên)

B. BÀI TẬP

1. Xâu con lặp dài nhất [REPS]

Cho xâu ký tự S . Hãy tìm xâu P dài nhất sao cho P xuất hiện trong S ít nhất k lần

Input:

- Dòng 1 chứa xâu S chỉ gồm các chữ cái in hoa, độ dài không quá 10^5
- Dòng 2 chứa số nguyên dương k .

Output: Xâu P tìm được (ghi ra xâu rỗng nếu không tồn tại xâu xuất hiện trong S ít nhất k lần).

Example:

input	output
MISSISSIPPI	ISSI
2	

2. Xâu đẩy vòng [STRINGS]

Xét xâu S độ dài n . Xâu các ký tự $S_k S_{k+1} \dots S_n S_1 S_2 \dots S_{k-1}$ được gọi là xâu đẩy vòng của S với $2 \leq k \leq n$. Chính S cũng chính là xâu đẩy vòng của S .

Xâu con liên tiếp của S là $S_i S_{i+1} \dots S_j$ với $1 \leq i \leq j \leq n$.

Yêu cầu: Cho 2 xâu a và b ($1 \leq \text{length}(b) \leq \text{length}(a) < 10^5$) chỉ gồm các ký tự thuộc tập $['a'..'z', 'A'..'Z', '0'..'9']$, hãy tính số lượng xâu con của a là xâu đẩy vòng của xâu b .

Input:

- Dòng 1: xâu a ,
- Dòng 2: xâu b .

Output: Một dòng chứa một số nguyên là số lượng xâu con của a là xâu đẩy vòng của xâu b .

Example:

input	output
abcabc abc	4
abcabc acb	0
aaaaaaa aa	6

3. Tiền tố dài nhất [SHIFT]

Bạn được cho hai xâu ký tự A và B có cùng độ dài. Cả hai xâu đều chứa n ký tự chữ cái tiếng Anh thường ('a'..'z'). Một phép đẩy là phép thực hiện xóa ký tự đầu tiên của dãy và chèn ký tự này vào cuối dãy. Ví dụ, đối với xâu 'abcd' nếu thực hiện đẩy một lần ta được 'bcda', thực hiện đẩy hai lần được 'cdab'.

Nhiệm vụ của bạn là thực hiện một số phép đẩy (có thể bằng không) trên xâu B sao cho độ dài của tiền tố chung giữa A và B là lớn nhất có thể.

Input:

- Dòng đầu tiên chứa số nguyên dương n
- Dòng thứ hai chứa xâu A và dòng thứ ba chứa xâu B .

Output: Một số nguyên duy nhất là số phép đẩy cần thực hiện. Nếu như có nhiều cách đẩy khác nhau cùng thỏa mãn thì chọn cách đẩy có số phép đẩy là nhỏ nhất.

Example:

input	output
5	3

ccadd bddcc	
----------------	--

Ghi chú:

- Subtask 1: $n \leq 5000$ [30%]
- Subtask 2: $n \leq 10^4$ [30%]
- Subtask 3: $n \leq 10^6$ [40%]

4. Quay xâu [STRCAT]

Cho xâu S độ dài N và M xâu độ dài bằng nhau w_1, w_2, \dots, w_M . Biết rằng tồn tại một cách ghép các xâu w_i thành xâu S , hãy chỉ ra cách ghép đó.

Input:

- Dòng 1: hai số nguyên N, M ($1 \leq N, M \leq 10^6; M \nmid N$)
- Dòng 2: xâu S
- Dòng 3 ... $M + 3$: dòng $i + 2$ ghi xâu w_i

Output: Dòng 1: M số nguyên là một hoán vị của $(1, 2, \dots, M)$ thể hiện thứ tự ghép tìm được

Example:

input	output
12 3 abcdefghijkl ijkl abcd efgh	2 3 1

5. Con đường gốm sứ [CERAMIC]

Sau khi bê tông hóa đê chống lụt, thành phố quyết định cho khảm lên tường bê tông của đê tranh ghép tạo bởi các mảnh gốm sứ lấy từ các lò gốm nổi tiếng trong nước. Toàn bộ con đê được chia thành n phần có độ rộng giống nhau, mỗi phần gọi là một lô. Mỗi bức tranh khảm trên đó đều phải có độ rộng giống nhau, tức là bao gồm một số như nhau các lô liên tiếp và toàn bộ tường phải được phủ kín tranh từ đầu đến cuối, mỗi lô phải được tạo màu chủ đạo (gọi là màu của lô) từ một loại gốm đặc trưng lấy từ một lò gốm nào đó trong nước, ví dụ gốm màu xanh Cô ban từ lò gốm Ánh Hồng Quảng Ninh, gốm da lươn – từ Bát Tràng Hà Nội, gốm mộc hồng nhật – từ Biên Hòa Đồng Nai, . . . Các loại gốm này được đánh số từ 1 đến 50 000.



Hướng dẫn viên du lịch giới thiệu với khách tham quan là có 2 nhóm nghệ nhân được giao việc tạo hình và khảm tranh. Với mỗi nhóm các bức tranh của đều được đặc trưng bởi dãy số (c_1, c_2, \dots, c_k) , trong đó k là độ rộng của tranh, c_i – màu của lô, $i = 1 \div k$, các bức tranh khác nhau có thể khác nhau ở trình tự xuất hiện màu của các lô, ví dụ với dãy số đặc trưng $(2, 6, 2, 9)$, trình tự màu trong tranh có thể là $(9, 2, 2, 6)$ hoặc $(6, 9, 2, 2)$ nhưng không thể là $(6, 9, 2, 3)$. Các bức tranh được ghép với nhau rất hài hòa và khách tham quan không nhận biết được sự chuyển tiếp từ tranh này sang tranh khác. Tuy vậy nhiều khách tham quan vẫn muốn biết có bao nhiêu bức tranh đã tạo ra và trong đó số bức tranh của mỗi nhóm là bao nhiêu.

Hãy xác định số lượng tranh có thể có và số lượng tranh mỗi nhóm đã làm.

Input:

- Dòng đầu tiên chứa một số nguyên n – số lượng lô của con đê ($2 \leq n \leq 10^5$),
- Dòng thứ 2 chứa n số nguyên a_1, a_2, \dots, a_n – màu của các lô ($1 \leq a_i \leq 50\,000, i = 1 \div n$).

Output: Dòng đầu tiên chứa số nguyên m – số lượng phương án khác nhau chia con đường thành các bức tranh, nếu không có cách phân chia để đảm bảo phân biệt tranh của đúng 2 nhóm thì đưa ra số -1. Nếu có cách phân biệt thì ở mỗi dòng tiếp theo đưa ra 3 số nguyên k , p và q – độ rộng bức tranh, số tranh do nhóm 1 thực hiện và số tranh do nhóm 2 thực hiện, thông tin đưa ra theo thứ tự tăng dần của k và ở mỗi dòng có $p \geq q > 0$.

Example:

input	output
9	1
1 2 3 6 4 9 3 1 2	3 2 1