



OLYMPIC TIN HỌC SINH VIÊN LẦN THỨ XXIV, 2015

Khởi thi: Chuyên tin

Thời gian làm bài: 180 phút

Ngày thi: 25-11-2015

Nơi thi: ĐẠI HỌC KINH DOANH VÀ CÔNG NGHỆ HÀ NỘI

TỔNG QUAN ĐỀ THI

Tên bài	Tên file chương trình	Tên file dữ liệu	Tên file kết quả	Hạn chế thời gian cho mỗi test
BOT	BOT.???	BOT.INP	BOT.OUT	0.3 giây
CON ĐƯỜNG GÓM SỨ	CERAMIC.???	CERAMIC.INP	CERAMIC.OUT	0.5 giây
THUẦN CHỪNG	PURE.???	PURE.INP	PURE.OUT	0.5 giây
TIẾN HÓA	EVOLUTION.???	EVOLUTION.INP	EVOLUTION.OUT	0.5 giây

Chú ý:

- Dấu ??? được thay thế bởi đuôi ngầm định của ngôn ngữ được sử dụng để cài đặt chương trình.
- Thí sinh phải nộp cả file mã nguồn của chương trình và file chương trình thực hiện (chương trình đã được biên dịch ra file .exe).

Hãy lập trình giải các bài sau đây:

Bài 1. BOT (20 điểm)

BOT (*Built-Operation-Transfer*, có nghĩa: Xây dựng-Vận hành-Chuyển giao) là hình thức Chính phủ kêu gọi các công ty bỏ vốn xây dựng trước (*Built*) thông qua đấu thầu, sau đó khai thác vận hành một thời gian (*Operation*) và sau cùng là chuyển giao (*Transfer*) lại cho nhà nước sở tại.

Đường cao tốc xuyên quốc gia được xây dựng theo hình thức BOT. Công ty Đa quốc gia *Modern Highway* trúng thầu, chia toàn bộ con đường thành n đoạn, đánh số từ 1 đến n . Theo tính toán của Công ty, cho đến khi chuyển giao con đường cho chính phủ sở tại quản lý thì lãi thu được ở đoạn đường thứ i là a_i , a_i có thể dương, âm hoặc bằng 0, tức là với từng đoạn con có thể lãi, lỗ hoặc hòa vốn. Từng nhóm các đoạn đường liên tiếp nhau (gọi tắt là khoảng) được chia cho các công ty con thực hiện. Công ty con *ASEAM Highway* hiện đang có trụ sở ở nước sở tại được quyền chọn trước khoảng tùy ý (có thể là cả con đường).

Dĩ nhiên Ban Giám đốc *ASEAM Highway* muốn chọn khoảng bắt đầu từ đoạn p đến hết đoạn q ($1 \leq p \leq q \leq n$) mang lại lợi nhuận cao nhất hoặc lỗ ít nhất nếu không có khoảng nào cho lãi.

Hãy chỉ ra khoảng cần chọn và lãi thu được. Nếu có nhiều cách chọn thì chỉ ra cách chọn có p nhỏ nhất.

Dữ liệu: Vào từ file văn bản BOT.INP:

Đầu tiên chứa số nguyên n ($1 \leq n \leq 10^6$),

✚ Dòng thứ 2 chứa n số nguyên a_1, a_2, \dots, a_n ($0 \leq |a_i| \leq 10^9, i = 1 \div n$).

Kết quả: Đưa ra file văn bản BOT.OUT trên một dòng 2 số nguyên p, q và lãi thu được.

Ví dụ:

BOT.INP															
16															
2	-4	5	-8	4	-1	-1	1	1	1	-2	2	4	-6	9	-4

BOT.OUT		
5	15	12

Bài 2. CON ĐƯỜNG GỐM SỨ (30 điểm)

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 đặc trưng (gọi là màu của lô) từ một loại gốm tiêu biểu 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. Mỗi bức tranh được biểu diễn bởi một dãy màu đặc trưng (c_1, c_2, \dots, c_k) , trong đó k là độ rộng của tranh, c_i – màu của lô, $i = 1 \div k$. Hai dãy màu đặc trưng được gọi là giống nhau nếu có thể đưa dãy này về dãy kia bằng phép hoán vị trình tự. Ví dụ dãy màu đặc trưng (1, 2, 5, 6) giống dãy màu đặc trưng (5, 6, 1, 2). Mỗi nhóm nghệ nhân có một dãy màu đặc trưng, và dãy màu đặc trưng của hai nhóm là không giống nhau. Hai bức tranh thuộc cùng một nhóm nghệ nhân thì hai dãy màu đặc trưng của chúng phải giống nhau. Dãy 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.

Dữ liệu: Vào từ file văn bản CERAMIC.INP:

✚ 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 ($1 \leq a_i \leq 50\,000, i = 1 \div n$).

Kết quả: Đưa ra file văn bản CERAMIC.OUT, 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 $p \geq q > 0$.

Ví dụ:

CERAMIC.INP
9
1 2 3 6 4 9 3 1 2

CERAMIC.OUT
1
3 2 1

Bài 3. THUẦN CHỦNG (30 điểm)

Gene là một đoạn kết gán các cặp ADN, mỗi cặp ADN được đặc trưng bằng một chữ cái trong tập $\{A, C, G, T\}$. Gene thuần chủng là gene hình thành từ một đoạn ADN cơ sở độ dài không quá m , được gán kết lặp đi lặp lại nhiều lần và ở lần lặp cuối cùng có thể chỉ chứa phần đầu của đoạn cơ sở. Gene được mô tả dưới dạng xâu S chỉ chứa các ký tự trong tập nêu trên. Như vậy gene thuần chủng là xâu có thể biểu diễn như tổng của k đoạn cơ sở ($k \geq 0$) và có thể có thêm một đoạn đầu của cơ sở.

Ví dụ, với $m = 10$, $S = \text{"ACATAGACATAGACATAGACA"}$ là một gene thuần chủng vì có đoạn cơ sở là "ACATAG" và $S = \text{"ACATAG"} + \text{"ACATAG"} + \text{"ACATAG"} + \text{"ACA"}$, nhưng với $m = 5$ thì S không phải là gene thuần chủng.

Cho gene S và giá trị m . Hãy xác định S có phải là gene thuần chủng hay không và đưa ra *đoạn cơ sở ngắn nhất* nếu S là gene thuần chủng hoặc đưa ra thông báo **"NO"** trong trường hợp ngược lại.

Dữ liệu: Vào từ file văn bản PURE.INP:

- ✚ Dòng đầu tiên chứa số nguyên m ($1 \leq m \leq 10^6$),
- ✚ Dòng thứ 2 chứa xâu S độ dài không quá 10^6 và chỉ chứa các ký tự trong tập đã nêu.

Kết quả: Đưa ra file văn bản PURE.OUT đoạn cơ sở ngắn nhất tìm được hoặc thông báo **NO**.

Ví dụ:

PURE.INP
10
ACATAGACATAGACATAGACA

PURE.OUT
ACATAG

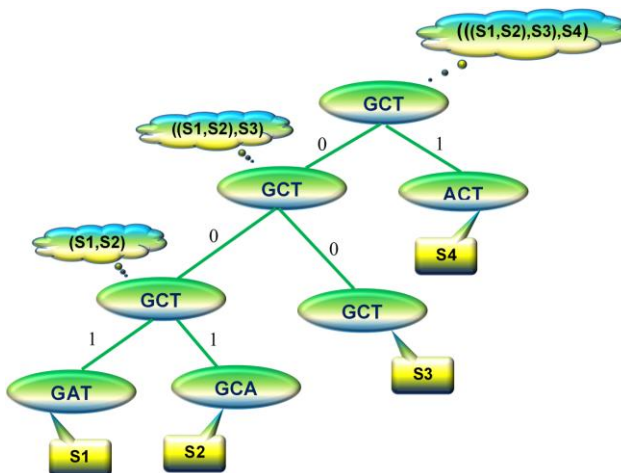
Bài 4. TIẾN HÓA (20 điểm)

Theo học thuyết tiến hóa của Darwin, các loài sinh vật tiến hóa từ một tổ tiên chung. Quá trình tiến hóa của các loài sinh vật có thể được xác định bằng cách phân tích các chuỗi protein của chúng. Chuỗi protein là một đoạn gán kết các axit amin. Có 20 loại axit amin khác nhau, mỗi axit amin được biểu diễn bởi một ký tự trong tập $\{A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V\}$. Độ dài của một chuỗi protein là số lượng các axit amin trong chuỗi đó.

Cho hai chuỗi protein $p = (p_1, p_2, \dots, p_k)$ và $q = (q_1, q_2, \dots, q_k)$ có cùng độ dài k , khoảng cách giữa hai chuỗi p và q được tính bằng số lượng vị trí i ($i = 1 \dots k$) mà p_i khác q_i . Ví dụ, khoảng cách giữa hai chuỗi protein $p=\text{GAT}$ và $q=\text{GCA}$ là 2.

Quá trình tiến hóa của các loài sinh vật có thể biểu diễn bởi một cây nhị phân có gốc. Giá trị mỗi nút của cây là chuỗi protein của sinh vật tương ứng với nút đó. Các nút lá chứa chuỗi protein của các sinh vật hiện tại, các nút trong của cây chứa chuỗi protein của các sinh vật tổ tiên. Nếu S_i và S_j là hai chuỗi protein của hai loài sinh vật có chung một tổ tiên trực tiếp, thì tổ tiên chung của chúng được biểu diễn bởi (S_i, S_j) . Biểu diễn của sinh vật tổ tiên tại nút gốc cho biết cấu trúc của cây (xem Hình bên dưới).

Giáo sư Haeseler muốn nghiên cứu quá trình tiến hóa của n loài sinh vật hiện tại, được biểu diễn bởi n chuỗi protein $S_1, S_2, S_3, \dots, S_n$. Các chuỗi protein đều có cùng độ dài là k . Một khó khăn trong quá trình nghiên cứu là Giáo sư không có thông tin về các chuỗi protein của các loài sinh vật tổ tiên nằm ở các nút bên trong của cây. Mục tiêu của Giáo sư là xác định chuỗi protein cho các loài sinh vật tổ tiên, để tổng độ dài các cạnh của cây tiến hóa là nhỏ nhất (độ dài cạnh nối hai loài sinh vật được tính bằng khoảng cách giữa hai chuỗi protein biểu diễn hai loài sinh vật đó).



Dữ liệu: Vào từ file EVOLUTION.INP:

- Dòng đầu tiên chứa hai số nguyên dương n và k ($2 \leq n < 100$; $1 \leq k \leq 1000$).
- Dòng thứ i trong số n dòng tiếp theo chứa chuỗi protein S_i có độ dài k biểu diễn cho sinh vật thứ i .
- Dòng cuối cùng chứa xâu biểu diễn của sinh vật tổ tiên ở nút gốc của cây.

Kết quả: Ghi ra file EVOLUTION.OUT một dòng duy nhất là tổng độ dài các cạnh của cây tiến hóa.

Ví dụ:

EVOLUTION.INP
4 3
GAT
GCA
GCT
ACT
(((S1, S2), S3), S4)

EVOLUTION.OUT
3

----- Hết -----