



# OLYMPIC TIN HỌC SINH VIÊN LẦN THỨ XVI, 2007

## Kh i thi: Cá nhân Cao  ng

Th i gian làm bài: 180 phút

Ngày thi: 10/11/2007

N-i thi: §<sup>1</sup> i hăc §<sub>μ</sub> N½NG

Tên bài	File ngu n n p	File d li u	File k t qu	Th i gian m i test
Tính l  ng	LUONG.XLS			
Xâu l p	WLOOP.*	WLOOP.INP	WLOOP.OUT	1 giây
Nh t kí hành trình	TRACE.*	TRACE.INP	TRACE.OUT	1 giây

Chú ý:

- D u \* c thay th b i uôi ng m nh c a ngôn ng c s d ng cài ch ng trình;
- Thí sinh ph i n p c file mã ngu n c a ch ng trình và file ch ng trình th c hi n (ch ng trình ã c biên d ch ra file .exe).

### Bài 1. Tính l  ng

N m 2007, l  ng tháng c a nhân viên Công ty Tỉnh Nguy n c tính d a trên s n m công tác, m c l  ng chính và x p lo i lao  ng.

- S n m công tác c tính b ng cách l y ph n nguyên c a t ng s ngày ã công tác t ngày vào công ty cho n ngày hôm nay (ngày 10 tháng 11 n m 2007) chia cho 365;
- Có 4 m c lao  ng c x p lo i: 1/2/3/4, ng v i m i m c có m t s tỉ n th ng, ph t t ng ng là 500000, 200000, 0, - 100000.
- L  ng th c t c a nhân viên là t ng c a l  ng chính, ph c p (3% l  ng chính nhân v i s n m công tác) và s tỉ n th ng ph t.

Hãy s d ng MS Excel t o t p **LUONG.XLS** gi i quy t m t s v i c v tính l  ng cho Công ty Tỉnh Nguy n, có 100 nhân viên.

Gi s trên **Sheet 1** d li u s c nh p vào các ô Ak, Bk, Ck t ng ng là ngày vào công ty, m c x p lo i lao  ng (1, 2, 3, ho c 4) và l  ng chính (không quá m i tri u) c a các nhân viên, v i k = 1, ..., 100. L p các công th c tính tỉ n th ng trung bình c a các nhân viên x p lo i 1 và 2 trong tháng, và t ng tỉ n l  ng th c t mà công ty ph i tr trong tháng cho toàn b nhân viên; và k t xu t t ng ng vào ô **G1** và **H1** c a **Sheet 1**.

Ch ng h n, v i s nhân viên là 5 ta có b ng m u sau:

	A	B	C	D	E	F	G	H
1	01/10/1971	1	10000000				300000.00	41355120.73
2	13/03/2001	2	2366669					
3	06/10/2007	4	540000					
4	11/12/2005	3	9136044					
5	19/10/1999	2	5655102					

Chú ý r ng b n ph i nh d ng ki u d li u cho các ô c t A là Date/Custom (khuôn d ng dd/mm/yyyy, ví d : 20/11/2007); c t B và C là Number không có s th p phân và không có d u tách nhóm s (ví d : 6668); ô **G1** và **H1** là Number v i 2 ch s sau d u ch m th p phân và không có d u tách nhóm s (ví d : 12345.68).

B n có th s d ng các ô ngoài 2 ô G1, H1 và các ô Ak, Bk, Ck, v i  $k = 1, \dots, 100$  t o các công th c trung gian.

**Ghi chú:** Bài này s c ch m b ng cách nh p d li u c a các test khác nhau vào các ô Ak, Bk, Ck, v i  $k = 1, \dots, 100$  và sau ó ki m tra k t qu các ô **G1** và **H1** trong *Sheet 1* c a t p **LUONG.XLS** mà thí sinh n p.

## Bài 2. Xâu l p

Trong gi th c hành so n th o v n b n trên MS Word, m i sinh viên c n gõ m t xâu kí t b t kì. Khi xem xét xâu k t qu c a sinh viên, giáo viên phát hi n ra r ng có m t s xâu l p c bi t c t o thành b ng cách ghép k l n m t xâu con ( $k > 1$ ). Ví d , 'ABABAB' là m t xâu c bi t, l p l i 3 l n xâu con 'AB'.

Hãy l p trình giúp giáo viên tìm s các xâu l p trong s  $N$  xâu mà các sinh viên ã gõ, bi t r ng các xâu này khác r ng, ôi m t khác nhau và dài c a m i xâu không quá 255.

**D li u:** vào t file v n b n **WLOOP.INP** có c u trúc nh sau:

- Dòng u tiên ch a s nguyên d ng  $N$ , v i  $N \leq 1000$ ;
- Dòng th  $i$  trong  $N$  dòng ti p theo ch a m t xâu mà sinh viên th  $i$  ã gõ, v i  $i = 1, \dots, N$ .

**K t qu :** ghi ra file v n b n **WLOOP.OUT** g m m t s nguyên không âm là s l ng xâu l p tìm c.

**Ví d :**

WLOOP.INP	WLOOP.OUT
4 Welcom!welcom! Welcom!Welcom! VOI2007&ACM/ICPC DANANG, 9-11/2007	1

## Bài 3. Nh t kí hành trình

oàn thám hi m sa m c Gô bi xu t phát t i m có t a  $(X_S, Y_S)$  sau  $N$  ngày kh o sát sa m c ã hoàn thành xu t s c các nhi m v ra và v t i ích an toàn i m có t a  $(X_D, Y_D)$ . u m i ngày trong cu c hành trình, khi m t tr i còn ch a k p t a ánh n ng chói chang nh mu n thiêu t m i sinh v t trên sa m c, oàn thám hi m di chuy n t i i m kh o sát m i, cách i m hi n t i m t n v dài và i theo m t trong s 4 h ng: ông (E), B c (N), Tây (W) ho c Nam (S). Nh v y, n u ban u tr c khi lên ng v trí c a oàn t a  $(X, Y)$  thì v trí m i n i oàn th c hi n các kh o sát và ng qua êm s là nh sau:

H ng i	T a X m i	T a Y m i
E	$X+1$	$Y$
N	$X$	$Y+1$
W	$X-1$	$Y$
S	$X$	$Y-1$

ng i c ghi l i trong nh t kí công tác d i d ng xâu kí t T ch ch a các kí t thu c t p {E, N, W, S}.

Ví d , t i m xu t phát  $X_S = 1, Y_S = 2$ , v i hành trình  $T = \text{'ENWNEEESESWSW'}$ , i m ích c a chuy n kh o sát s là  $X_D = 2, Y_D = 1$ .

Từ theo yêu c u th c t , m t i m có th c quay l i kh o sát nhi u l n.

M i vi c d ng nh vô cùng t t p n u nh không có m t s c nh : báo cáo công tác (in trên máy vi tính) b tr v b sung ch nh lý vì trong xâu T xác nh hành trình có 2 kí t l i n nhau không ch h ng i, t c là không thu c t p {E, N, W, S}! Ví d , hành trình trên có th b gõ nh m vào máy vi tính thành  $T = \text{'ENWNEYZSESWSW'}$ .

**Yêu c u:** hãy giúp oàn thám hi m s a l i các kí t sai trong báo cáo hành trình. N u có nhi u cách s a h p lý thì ch ra cách s a cho xâu T nh nh t theo th t t i n.

**D li u:** vào t file v n b n **TRACE.INP** g m 2 dòng:

- Dòng th nh t ghi 4 s nguyên  $X_S, Y_S, X_D, Y_D$ , có giá tr tuy t i không quá  $10^6$ , gi a các s cách nhau m t d u cách.
- Dòng th hai ghi xâu T có dài không quá 100 ch a 2 kí t l i n nhau.

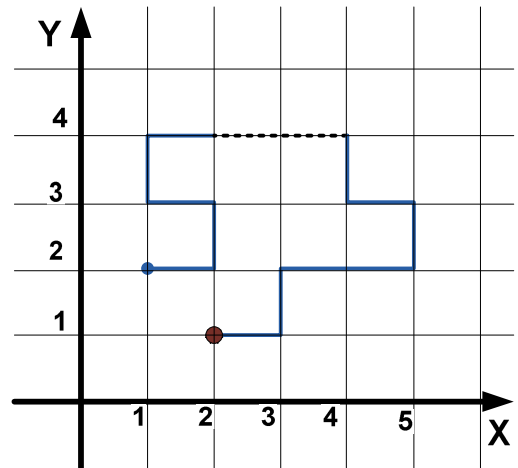
D li u m b o test có l i gi i.

**K t qu :** ghi ra file v n b n **TRACE.OUT** có m t dòng ch a xâu T ã ch nh lý.

**Ví d :**

TRACE.INP
1 2 2 1
ENWNEYZSESWSW

TRACE.OUT
ENWNEEESESWSW



$T = \text{'ENWNEEESESWSW'}$

----- H t -----



# OLYMPIC TIN HỌC SINH VIÊN LẦN THỨ XVI, 2007

## Kh i thi: Cá nhân Không chuyên

Th i gian làm bài: 180 phút

Ngày thi: 10/11/2007

N-i thi: §<sup>1</sup> i hãc §µ N½NG

Tên bài	File ngu n n p	File d li u	File k t qu	Th i gian m i test
Tính l ãng	LUONG.XLS			
Soát vé	TICKET.*	TICKET.INP	TICKET.OUT	1 giây
Nh t kí hành trình	TRACE.*	TRACE.INP	TRACE.OUT	1 giây

### Chú ý:

- D u \* c thay th b i uôi ng m nh c a ngôn ng c s d ng cài ch ãng trình;*
- Thí sinh ph i n p c file mã ngu n c a ch ãng trình và file ch ãng trình th c hi ãn (ch ãng trình ã c biên d ch ra file .exe).*

### Bài 1. Tính l ãng

N m 2007, l ãng tháng c a nhân viên Công ty Tỉnh Nguy ãn c tính d a trên s ãn m công tác, m c l ãng chính và x p lo i lao ãng.

- S ãn m công tác c tính b ãng cách l y ph ãn nguyên c a t ãng s ãng ã công tác t ãng vào công ty cho ãn ngày hôm nay (ngày 10 tháng 11 n m 2007) chia cho 365;
- Có 4 m c lao ãng c x p lo i: 1/2/3/4, ãng v i m i m c có m t s ãn th ãng, ph ãt ãng ãng là 500000, 200000, 0, - 100000.
- L ãng th c t c a nhân viên là t ãng c a l ãng chính, ph ãc p (3% l ãng chính nhân v i s ãn m công tác) và s ãn th ãng ph ãt.

Hãy s d ãng MS Excel t o t p **LUONG.XLS** gi i quy t m t s v i c v tính l ãng cho Công ty Tỉnh Nguy ãn, có 100 nhân viên.

Gi s ãn trên **Sheet 1** d li u s c nh p vào các ô Ak, Bk, Ck t ãng ãng là ngày vào công ty, m c x p lo i lao ãng (1, 2, 3, ho c 4) và l ãng chính (không quá m i tr i u) c a các nhân viên, v i k = 1, ..., 100. L p các công th c tính t i ãn th ãng trung bình c a các nhân viên x p lo i 1 và 2 trong tháng, và t ãng t i ãn l ãng th c t mà công ty ph i tr ãn trong tháng cho toàn b ãn nhân viên; và k t xu t t ãng ãng vào ô **G1** và **H1** c a **Sheet 1**.

Ch ãng h ãn, v i s ãn nhân viên là 5 ta có b ãng m u sau:

	A	B	C	D	E	F	G	H
1	01/10/1971	1	10000000				300000.00	41355120.73
2	13/03/2001	2	2366669					
3	06/10/2007	4	540000					
4	11/12/2005	3	9136044					
5	19/10/1999	2	5655102					

Chú ý r ng b n ph i nh d ng ki u d li u cho các ô c t A là Date/Custom (khuôn d ng dd/mm/yyyy, ví d : 20/11/2007); c t B và C là Number không có s th p phân và không có d u tách nhóm s (ví d : 6668); ô **G1** và **H1** là Number v i 2 ch s sau d u ch m th p phân và không có d u tách nhóm s (ví d : 12345.68).

B n có th s d ng các ô ngoài 2 ô G1, H1 và các ô Ak, Bk, Ck, v i  $k = 1, \dots, 100$  t o các công th c trung gian.

**Ghi chú:** Bài này s c ch m b ng cách nh p d li u c a các test khác nhau vào các ô Ak, Bk, Ck, v i  $k = 1, \dots, 100$  và sau ó ki m tra k t qu các ô **G1** và **H1** trong **Sheet 1** c a t p **LUONG.XLS** mà thí sinh n p.

## Bài 2. Soát vé

M t xe buýt ho t ng trên tuy n ng có  $N$  b n và các b n c ánh s t l n  $N$ , (1 là b n xu t phát,  $N$  là b n cu i cùng). b n  $i$  có  $A_i$  hành khách lên xe và  $B_i$  hành khách xu ng xe, v i  $i = 1, \dots, N$ . V i s nguyên d ng  $K$  cho tr c, v i c soát vé hành khách c nhân viên th c hi n các b n :  $1, K+1, 2K+1, \dots, mK+1$ , trong ó  $mK+1 \leq N$  và theo quy t c là sau khi các hành khách ã lên và xu ng xe xong thì soát vé t t c các hành khách có m t trên xe. V i cách soát vé nh v y, m t s hành khách có th không b soát vé l n nào.

**Yêu c u:** hãy l p trình tìm s hành khách ít nh t và nhi u nh t i trên xe buýt mà không b soát vé l n nào.

**D li u:** vào t file v n b n **TICKET.INP** có d ng:

- Dòng th nh t ch a 2 s nguyên d ng  $N$  và  $K$ , v i  $1 \leq K \leq N \leq 1000$ ;
- Dòng th  $i$  trong  $N$  dòng ti p theo ch a 2 s nguyên  $A_i$  và  $B_i$ , v i  $0 \leq A_i, B_i \leq 1000$ ,  $i = 1, \dots, N$ .

Các s trên cùng m t dòng cách nhau ít nh t m t d u cách.

**K t qu :** ghi ra file v n b n **TICKET.OUT** g m m t dòng ch a 2 s nguyên, cách nhau m t d u cách, là s hành khách nh nh t và l n nh t mà không b soát vé l n nào.

**Ví d :**

TICKET.INP	TICKET.OUT
4 2	0 3
5 0	
5 0	
0 3	
0 7	

## Bài 3. Nh t kí hành trình

oàn thám hi m sa m c Gô bi xu t phát t i m có t a  $(X_S, Y_S)$  sau  $N$  ngày kh o sát sa m c ã hoàn thành xu t s c các nhi m v ra và v t i ích an toàn i m có t a  $(X_D, Y_D)$ . u m i ngày trong cu c hành trình, khi m t tr i còn ch a k p t a ánh n ng chói chang nh mu n thiêu t m i sinh v t trên sa m c, oàn thám hi m di chuy n t i i m kh o sát m i, cách i m hi n t i m t n v dài và i theo m t trong s 4 h ng: ông (E), B c (N), Tây (W) ho c Nam (S). Nh v y, n u ban u tr c khi lên ng v trí c a oàn t a  $(X, Y)$  thì v trí m i n i oàn th c hi n các kh o sát và ng qua êm s là nh sau:

Hướng	Tọa độ mới	Tọa độ cũ
E	$X+1$	$Y$
N	$X$	$Y+1$
W	$X-1$	$Y$
S	$X$	$Y-1$

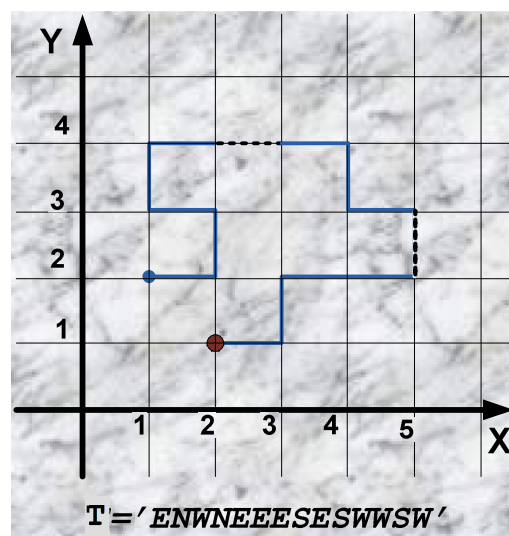
Hướng đi được ghi lại trong nhát kí công tác để đi đúng hướng kí tự T chứa các kí tự thuộc tập {E, N, W, S}.

Ví dụ, điểm xuất phát  $X_S = 1, Y_S = 2$ , và hành trình  $T = \text{'ENWNEEESESWSW'}$ , điểm đích của chuyển khảo sát sẽ là  $X_D = 2, Y_D = 1$ .

Tùy theo yêu cầu thực tế, điểm đích có thể quay lại khảo sát ngẫu nhiên.

Một ví dụ đáng nhẽ vô cùng thú vị nhưng không có mặt sự kiện: báo cáo công tác (in trên máy vi tính) bắt buộc bổ sung chênh lệch vì trong chuỗi T xác định hành trình có mặt sự kiện 1 không chênh hướng, tức là không thuộc tập {E, N, W, S}! Ví dụ, hành trình trên có thể bỏ nhậm vào máy vi tính thành  $T = \text{'ENWNEYESEZWSW'}$ .

Cần phải liệt kê các kí tự sai nên chuỗi mới bắt đầu từ điểm xuất phát  $(X_S, Y_S)$  hành trình sẽ khác thức điểm  $(X_D, Y_D)$ . Hai cách sẽ có giá trị khác nhau nếu cho các chuỗi mô tả hành trình khác nhau.



**Yêu cầu:** hãy giúp xác định có bao nhiêu cách sắp các kí tự sai có mặt chuỗi hành trình ứng và đưa ra chuỗi T có thể tính toán được.

**Dữ liệu:** vào tệp file văn bản **TRACE.INP** gồm 2 dòng:

- Dòng thứ nhất ghi 4 số nguyên  $X_S, Y_S, X_D, Y_D$ , có giá trị tuyệt đối không quá  $10^6$ , giữa các số cách nhau một dấu cách.
- Dòng thứ hai ghi chuỗi T có độ dài không quá 255 chứa không quá 10 kí tự sai.

Dữ liệu mẫu test có lợi ích.

**Kết quả:** ghi ra file văn bản **TRACE.OUT**:

- Dòng thứ nhất chứa mặt số nguyên là số cách sắp;
- Dòng thứ hai chứa chuỗi T đã chỉnh sửa có thể tính toán được.

**Ví dụ:**

TRACE.INP
1 2 2 1
ENWNEYESZWSW

TRACE.OUT
2
ENWNEEESESWSW

**Lưu ý:** có không ít hơn 60% số test với độ dài chuỗi T không quá 100 và số kí tự sai không quá 3. Giải đúng các test này, thí sinh sẽ không ít hơn 60% số điểm tối đa cho toàn bộ bài toán.

----- H t -----



## **Problems Overview**

**Problem A: Digit Counting**

**Problem B: Numerical surprises**

**Problem C: Prime k-tuple**

**Problem D: The longest constant gene**

**Problem E: Lazy Susan**

**Problem F: Square Painting**

**Problem G: The Ultimate Password**

**Problem H: Plagiarism Detection**

**Problem I: Space Beacon**

**Problem J: Ig-Nobel Prize**

*Note: The input and output for all the problems are standard input and output.*



## **Problem A: Digit Counting**

Trung is bored with his mathematics homeworks. He takes a piece of chalk and starts writing a sequence of consecutive integers starting with 1 to  $N$  ( $1 < N < 10000$ ). After that, he counts the number of times each digit (0 to 9) appears in the sequence. For example, with  $N = 13$ , the sequence is:

12345678910111213

In this sequence, 0 appears once, 1 appears 6 times, 2 appears 2 times, 3 appears 3 times, and each digit from 4 to 9 appears once.

After playing for a while, Trung gets bored again. He now wants to write a program to do this for him. Your task is to help him with writing this program.

### **Input**

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each test case, there is one single line containing the number  $N$ .

### **Output**

For each test case, write sequentially in one line the number of digit 0, 1, ...9 separated by a space.

<b>Sample Input</b>	<b>Sample Output</b>
2	0 1 1 1 0 0 0 0 0 0
3	1 6 2 2 1 1 1 1 1 1
13	





## **Problem B: Numerical surprises**

We suspect that for every positive integer  $N$  there exists an integer of the form  $11\dots10\dots0$  (a sequence of 1's followed by 0 or more 0's) that is divisible by  $N$ . For example, with  $N = 3$ , 111 is divisible by 3, with  $N = 4$ , 100 is divisible by 4, with  $N = 7$ , 11111 is divisible by 7. We want to verify this for some integers. The solution to this problem is to find two different numbers  $P$  and  $Q$  in the form of  $11\dots1$  (a sequence of 1's) that have the same remainder when dividing by  $N$ . The difference  $D$  between  $P$  and  $Q$  will be in the form of  $11\dots10\dots0$  and divisible by  $N$ .

In order to solve this problem, we have to start with finding the remainder when dividing a number in the form of  $11\dots1$  by  $N$ . Your task is to write a program to do this.

### **Input**

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

Each data set is described by two lines. The first line contains the integer  $N$  ( $1 < N < 10^9$ ). The second line contains the integer number  $P$  ( $P$  contains at least one digit and at most 1000 digits).

### **Output**

For each test case, write in one line the remainder when dividing  $P$  by  $N$ .

Sample Input	Sample Output
2	3
4	1
11	
5	
111	



## Problem C: Prime k-tuple

$\{p_1, \dots, p_k : p_1 < p_2 < \dots < p_k\}$  is called a prime  $k$ -tuple of distance  $s$  if  $p_1, p_2, \dots, p_k$  are consecutive prime numbers and  $p_k - p_1 = s$ . For example, with  $k = 4$ ,  $s = 8$ ,  $\{11, 13, 17, 19\}$  is a prime 4-tuple of distance 8.

Given an interval  $[a, b]$ ,  $k$ , and  $s$ , your task is to write a program to find the number of prime  $k$ -tuples of distance  $s$  in the interval  $[a, b]$ .

### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each data set, there is only one line containing 4 numbers,  $a$ ,  $b$ ,  $k$  and  $s$  ( $a, b < 2 \cdot 10^9$ ,  $k < 10$ ,  $s < 40$ ).

### Output

For each test case, write in one line the numbers of prime  $k$ -tuples of distance  $s$ .

Sample Input	Sample Output
1 100 200 4 8	2

## Problem D: The longest constant gene

Thousands of genomes, from virus to human, are available in public databases. Each genome is presented as a string of nucleotides: “A”, “C”, “G”, and “T”. To study the relationship among organisms, their genomes are analyzed. A sequence of nucleotides is called a *constant gene* if it appears in all genomes.

Given  $N$  genomes, your task is to write a program to find the longest constant gene among these genomes.

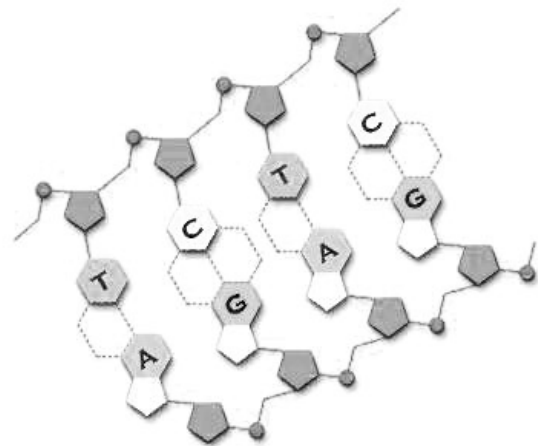
### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each data set, the first line contains the integer  $N$  ( $1 < N < 7$ ) indicating the number of genomes. Each line in the next  $N$  following lines contains one genome (the length of each genome is limited to one million).

### Output

For each test case, write in one line an integer number indicating the length of the longest constant genome.

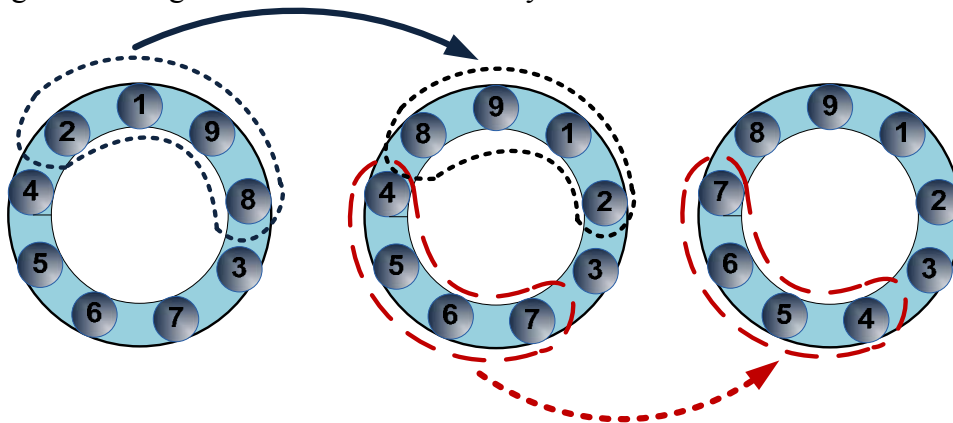


Sample Input	Sample Output
2 2 ACGGGCGTCGTCCCCGTCGTTCGTATC CTCGTCGTCCCCGTCGTTCGTGTC	18 11
3 ACGACGGCTGCGGTAAACC TTACGGCTGCGGTCCCCTT CCCCCGTTTACGGCTGCGGTGG	

## Problem E: Lazy Susan

There are  $N$  marbles, which are labeled  $1, 2, \dots, N$ . The  $N$  marbles are put in a circular track in an arbitrary order. In the top part of the track there is a “lazy Susan”, which is a tray that can hold exactly 4 marbles. The tray can be rotated, reversing the orientation of the four marbles. The tray can also be moved around the track in both directions.

For example, 9 marbles  $1, 9, 8, 3, 7, 6, 5, 4, 2$  are put in the circular track in clockwise order as shown in the following figure. This figure also shows how the tray is moved and rotated.



Trung wants you to arrange the marbles by moving and rotating the tray so that when listing the marbles from some position in the track in clockwise order, we get  $(1, 2, \dots, N)$ . Your task is to write a program to tell Trung that either this can be done or not.

### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 100. The following lines describe the data sets.

For each data set, the first line contains the integer  $N$  ( $8 \leq N \leq 500$ ). The second line describes the initial state of the track. It contains  $N$  numbers which are the labels of the marbles when listing in clockwise order.

### Output

For each test case, write in one line “**possible**” if there exists a solution to arrange the marbles. If not so, write “**impossible**”.

Sample Input	Sample Output
2	possible
9	impossible
1 9 8 3 7 6 5 4 2	
11	
1 3 2 4 5 6 7 8 9 10 11	



## Problem F: Square Painting

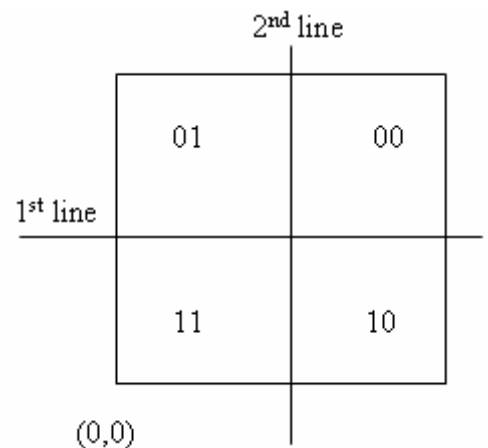
Given a  $10000 \times 10000$ -square in the Cartesian plane with 4 corners  $(0,0)$   $(0,10000)$   $(10000,10000)$   $(10000,0)$ . We consecutively draw  $N$  lines, which are numbered from 1 to  $N$ . These lines subdivide the given square into  $M$  pieces, of which all edges are longer than 1. Suppose that each line intersects exactly two edges of the given square and each line does not pass any square corner.

The  $i^{\text{th}}$  line divides the square into two parts:  $P_i$  containing  $(0,0)$ , and  $Q_i$  not containing  $(0,0)$ . After all the subdivisions, each resulted piece is labeled by a nonnegative integer, which has binary representation  $b_1b_2\dots b_n$ , where  $b_i=1$  if the piece lies in  $P_i$ , and  $b_i=0$  if the piece lies in  $Q_i$ . For example, the two lines in the figure subdivide the square into 4 pieces.

Now they want to paint all the pieces with  $S$  colors denoted by integers from 1 to  $S$  such that:

- Each piece is painted with only one color;
- Two adjacent pieces (having a common edge) must be painted with two different colors.
- The piece with the smallest label must be painted with color 1.

Your task is to write a program that determines the solution to paint the pieces with the minimum number of colors  $S$ .



### Input

The input consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each data set, the first line contains a single positive integer  $N$  ( $N \leq 30$ ) – the number of lines. The  $i^{\text{th}}$  line of the following  $N$  lines contains 4 **integers**  $x_i$ ,  $y_i$ ,  $z_i$ ,  $t_i$  – the coordinates of the intersections of  $i^{\text{th}}$  line with two square edges.

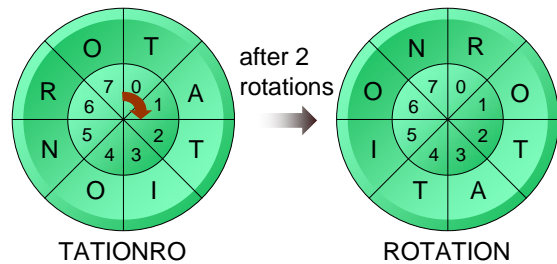
### Output

For each test case, output the solution in a number of lines. In the first line, write the minimum number of colors  $S$  needed for painting. The next  $M$  lines described how each piece is painted in increasing order of the label of all the pieces. In each line, write two numbers separated by a space, the former is the label of the piece and the latter is the color used to paint that piece.

Sample Input	Sample Output
2	2
1	0 1
0 5000 10000 5000	1 2
2	2
0 5000 10000 5000	0 1
5000 0 5000 10000	1 2
	2 2
	3 1

## Problem G: The Ultimate Password

A “letter” lock is a circle, in which we mark some positions equidistant from one another. The positions are numbered clockwise by a zero-based index and there is exactly one English capital letter put in each position. The state of the lock is given by a string, which contains all letters enumerated from position 0 to the end. The lock can change its state by performing rotations: After a rotation, every letter in the circle will move clockwise to the next position. If the lock received a password in form of a string, it will step by step rotate to change its state and verify whether its current state matches the prefix of the password. Whenever it matches, the lock will be unlocked, and the verification process will stop successfully. If the lock moves around and there has been no matched state found, it will delete the first character in the password and retry with the new password, and so on. This process repeats until the lock is unlocked or the password has been completely deleted (verification fails).



Given a sequence of  $N$  locks, in which the state of any lock is not shorter than the state of its previous lock, one may want to unlock all of them in succession by only one password. The process is as follows: the password is first applied to the first lock in the sequence; the remaining password after unlocking the first lock will be applied again to the second lock, the remaining password after unlocking the second lock will be applied again to the third lock, etc. The process continues until the last lock is unlocked.

Your task is to write a program to find the shortest password (in terms of the number of characters) that can be used to unlock all these  $N$  locks in succession. If there is more than one solution, just find the first one in lexicographic order.

### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each data set, there is only one line containing a string that lists all states of the locks separated by one semicolon (“;”). In each test case, the number of locks is not greater than 200, the state of every lock is not empty and consists of less than 101 letters.

### Output

For each test case, write in one line the shortest password to unlock all the locks in succession.

Sample Input	Sample Output
3 TOPO;POFTTO;THEPOF;HEWOOF;HEWORLD WELC;COMEEL;METLCO;TOCOME HEACMT;PROGRAMCM;RAMMINGCON;CONTESTING	TOPOFTHEWORLD WELCOMETO THEACMPROGRAMMINGCONTEST



## Problem H: Plagiarism Detection

A local school has noticed an increasing number of cases where electronic essays submitted by students are very similar. The teachers come up with an effective method to automatically detect similar documents using  $n$ -grams as the building block. An  $n$ -gram is a sequence of  $n$  consecutive words. Two  $n$ -grams are considered equal if their contents are exactly the same (case sensitive).

A word is a sequence of consecutive characters delimited by one or more whitespaces, punctuations or boundaries of email addresses. A whitespace can take the following ASCII codes: 9 (horizontal tab), 10 (new line), 11 (vertical tab), 12 (new page), 13 (carriage return) or 32 (space). A punctuation character can be any character with the following ASCII codes: 33 (exclamation mark), 44 (comma), 46 (period), 58 (colon), 59 (semicolon) and 63 (question mark). But when the period is part of an email address, the whole email address is considered to be a word.

An email address has the format: *local-part@domain-part*. Both the local part and the domain part may only contain the following ASCII characters:

- Uppercase and lowercase letters
- The digits 0 through 9
- The characters with ASCII codes 45(-) or 95(\_)
- The period (.) provided that it is not the first nor the last character in the local part or the domain part, nor may it appear two or more times consecutively.

Additionally, the domain part must contain at least one period and an email address is the longest string of characters satisfying the above-mentioned conditions.

For example, the following document:

*Email address, Email address:* [admin.team@company.com.vn](mailto:admin.team@company.com.vn).

has three unique 2-grams namely: “Email address”, “address Email” and “address [admin.team@company.com.vn](mailto:admin.team@company.com.vn)”. Notice that this document has only one email address, which is [admin.team@company.com.vn](mailto:admin.team@company.com.vn), and its substring [team@company.com](mailto:team@company.com) is not a valid email address.

For a given value of  $n$ , suppose document  $d_1$  has  $m$  unique  $n$ -grams, document  $d_2$  has  $l$  unique  $n$ -grams and the two documents have  $s$  unique  $n$ -grams in common. The similarity between two documents is calculated with the following formula:

$$\text{similarity}(d_1, d_2) = \min\left(\frac{s}{m}, \frac{s}{l}\right) * 100$$

Your task is to write a program to compute the similarity percentage between two given documents for a given value of  $n$ .

### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

# The ACM-ICPC 2007 Asia Regional Contest, Site Danang



For each data set, the first line of each data set contains one integer  $n$  ( $1 \leq n \leq 10$ ) where  $n$  denotes which  $n$ -gram to use. The second line contains two integers  $k$  and  $l$ , which are the number of lines the first and second documents contain respectively. The next  $k$  lines contain the content of the first document. The next  $l$  lines contain the content of the second document.

## Output

For each test case, write in one line the similarity percentage, truncated to 2 decimal places, between the two input documents.

Sample Input	Sample Output
1 2 2 1 John, my friend, is a good programmer. My friend John is a good programmer!	50.00





## Problem I: Space Beacon

In order to discover all the planets of the solar system, we want to develop techniques to travel safely through an asteroid belt between Mars and Jupiter. We plan to drop automatic-signaling devices into large asteroids of the belt, which will act as space beacons to guide the ships. They will assist autopilots to track the location of the ships to adjust the orbit. Each signal sent by each beacon contains a sequence of pulses, and is characterized by a sequence  $T$ :

$$T = t_1, t_2, \dots, t_k, (3 \leq k \leq 18),$$

where  $t_i$  is the duration of  $i$ -th pulse ( $t_i$  is integer and is in the range of 1..9).

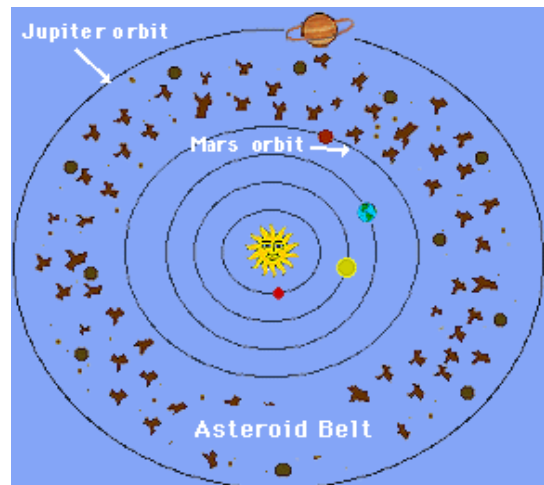
In order to simplify the technical checking process and to increase the signal recognition ability, the sequence  $T$  of each beacon is designed with the following criteria:

With  $1 < i < k$ , either:

$$\begin{cases} t_{i-1} < t_i, t_i > t_{i+1} & \text{for } i \bmod 2 = 0 \\ t_{i-1} > t_i, t_i < t_{i+1} & \text{for } i \bmod 2 = 1 \end{cases}$$

or

$$\begin{cases} t_{i-1} > t_i, t_i < t_{i+1} & \text{for } i \bmod 2 = 0 \\ t_{i-1} < t_i, t_i > t_{i+1} & \text{for } i \bmod 2 = 1 \end{cases}$$



All possible  $T$  sequences are sorted in lexicographic order and labeled by consecutive integers starting with 1. The label of the sequence  $T$  of each beacon is used as the identifier of the beacon.

Given the sequence  $T$  of a beacon, your task is to write a program to find the identifier of that beacon.

### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each data set, there is only one single line containing  $k$  integers  $t_1, t_2, \dots, t_k$  separated by space describing the  $T$  sequence of a beacon.

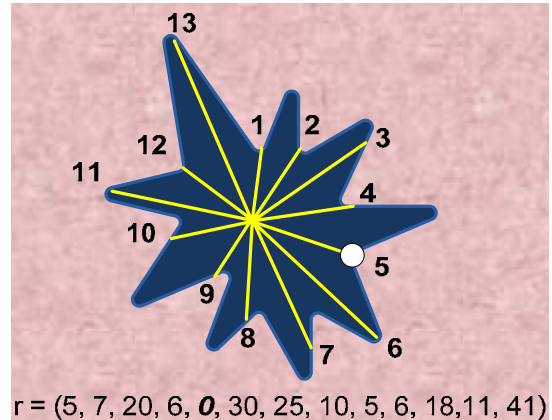
### Output

For each test case, write in one line the ID of the beacon with the given  $T$  sequence.

Sample Input	Sample Output
2	2
1 2 1 2	4
1 2 1 2 1 2	

## Problem J: Ig-Nobel Prize

One of the scientific research projects was almost chosen to the shortlist of the Ig Nobel. It was rejected because the author stated that the analyzed data is not sufficient, and that it will be studied further in the future. The research is to investigate the shape of ink drops left by grade 1 students in their notebooks. The author chose a point in the ink drop to act as the center. A line starting from that point rotates  $360^\circ$ ; and whenever it hits an angle of  $360^\circ/n$ , the intersection of the line and the border of the drop will be marked. At the  $i^{\text{th}}$  mark, we get the  $i^{\text{th}}$  intersection ( $i = 1, 2, \dots, n$ ). Thus, after the  $(n-1)^{\text{th}}$  is the  $n^{\text{th}}$ , and after the  $n^{\text{th}}$  is the  $1^{\text{st}}$  one. The author recorded the length  $r_i$  from the center to the  $i^{\text{th}}$  intersection by  $(1 \leq r_i \leq 10^6)$ . However, at the  $m^{\text{th}}$  marking of the intersection, he accidentally made a hole in the paper. Therefore, he didn't have the exact  $r_m$ . With the honesty of a scientist, he left  $r_m$  blank, recorded as a value of 0. One of the research goals is to find a rain drop which has a shape closest to that of the ink drop. In order to find it, the author planned to eliminate a minimal number of intersections so that the  $r_i$  of the remaining intersections can form a **descending order** of values starting from a certain intersection by either directions (in a circular way). In this process, the unknown value  $r_m$  is assigned by a value  $k$  in the range from 1 to  $n$ .



Given  $n$  and  $r_i$  ( $i = 1 \div n$ ), your task is to write a program to find the value of  $k$  in order that the number of the eliminated intersections is minimal. In the case that there is more than one value of  $k$ , find the smallest one.

### Input

The input file consists of several data sets. The first line of the input file contains the number of data sets which is a positive integer and is not bigger than 20. The following lines describe the data sets.

For each test case, the first line contains the integer  $n$  ( $1 < n \leq 10\,000$ ). The second line contains  $n$  integers  $r_1, r_2, \dots, r_n$  separated by space.

### Output

For each test case, write in one line two integers:  $k$  and the minimal number of eliminated intersections separated by one single space.

Sample Input	Sample Output
1 5 5 7 0 7 4	1 1