Data Structure and Algorithm, Spring 2018

Homework 1

Release: Tuesday, March 13, 2018

Due: 23:59:59, Sunday, March 25, 2018

TA E-mail: `dsa1@csie.ntu.edu.tw`

Rules and Instructions

- In homework 1, the problem set contains 5 problems and is divided into two parts, non-programming part (problem 1, 2, 3) and programming part (problem 4, 5).
- Please go to *DSA Judge* (`https://dsa.csie.org`) and complete the first problem to get familiar with the usage of the judge system.
- For problems in non-programming part, you should combine your solutions in ONE pdf file, and then hand it in via the judge system. Your solution must be as simple as possible. At the TAs' discretion, too complicated solutions will be counted as wrong.
- For problems in programming part, you should write your code in C programming language, and then hand them in via the judge system. You have 5 chances per day for each problem. The compilation command of the judge system will be `gcc main.c -static -std=c11 -O2 -lm`.
- Discussions with others are encouraged. However, you should write down your solutions in your own words. In addition, for each problem, you have to specify the references (the Internet URL you consulted with or the people you discussed with) on the first page of your solution to that problem.
- The score of the part that is submitted after the deadline will get some penalties according to the following rule:

$$LateScore = (\frac{86400 - DelayTime(sec.)}{86400})OriginalScore$$

Non-Programming Part

Problem 1. Stack/Queue (15 *points*)

You have learned in class that stack is a FILO/LIFO data structure and that queue is a FIFO/LILO data structure. Please describe a procedure under the following scenarios. They are all independent. You can assume that all data structure mentioned below have infinite capacity and all operations are valid. That is, no `Pop()` will be called on an empty stack and no `DeQueue()` will be called on an empty queue.

a. (4 *points*) Please use 2 queues and $O(1)$ extra space to simulate a stack. That is, please implement `Pop()`, `Push()` and `IsEmpty()` of a stack.

b. (4 *points*) Please use 2 stacks and $O(1)$ extra space to simulate a queue. That is, please implement `DeQueue()`, `EnQueue()` and `IsEmpty()` of a queue.

You are given a stack S and a queue Q. They both have infinite capacity, and they both store positive integers. The numbers of elements of S and Q will both be not greater than n. Please describe a procedure to determine whether there exists common element between S and Q in the following scenarios. If there exists, please output arbitrarily one of the common elements. If not, please output 0. The scenarios are independent. For each problem, if it's not trivial, please briefly explain why your procedure is correct and why the time/space complexity meets the constraints.

c. (2 *points*) Guaranteed that the elements in S, from bottom to top, are strictly increasing, and the elements in Q, from rear to front, are also strictly **increasing**. In $O(n)$ time, $O(1)$ extra space.

d. (5 *points*) Guaranteed that the elements in S, from bottom to top, are strictly increasing, while the elements in Q, from rear to front, are strictly **decreasing**. In $O(n)$ time, $O(1)$ extra space.

Problem 2. Complexity (20 *points*)

In this part, if you would like to use any theorem which is not mentioned in class, please prove it in advance.

$$(\lg(n) = \log_2(n),\ \ln(n) = \log_e(n),\ \log(n) = \log_{10}(n))$$

1. (1 *points*) Please rank the following functions by the order of growth. No proof or extra statement is needed.

$$4^n, 8^n, n^2, n^n, n\lg n, e^{\ln(n)}, 2^{\sqrt{\lg n}}$$

2. (11 *points*) Prove or disprove the following statements. You should provide a formal proof or a counterexample for each statement. Please note that in the following statements, $f(n), g(n), i(n), j(n)$ are *non-negative*, *monotonically increasing* functions.
   non-negative: The ranges of these functions are all non-negative.
   monotonically increasing:
   If $n_1 > n_2$, then $f(n_1) \geq f(n_2), g(n_1) \geq g(n_2), i(n_1) \geq i(n_2), j(n_1) \geq j(n_2)$.

   a. (1 *points*) If $f(n) = O(i(n))$, $g(n) = O(j(n))$, then $f(n) - g(n) = O(i(n) - j(n))$.

   b. (2 *points*) If $f(n) = O(g(n))$, then $f^2(n) = O(g^2(n))$.

   c. (2 *points*) If $f(n) = \Omega(i(n))$, $g(n) = \Omega(j(n))$, then $f(n) + g(n) = \Omega(i(n) + j(n))$

   d. (2 *points*) If $f(n) = O(g(n))$, then $2^{f(n)} = O(2^{g(n)})$

   e. (2 *points*) If $2^{f(n)} = O(2^{g(n)})$, then $f(n) = O(g(n))$

   f. (2 *points*) $f(n) + g(n) = \Theta(\max(f(n), g(n)))$

3. a. (4 *points*) Define a function $f(n)$ as below:

$$f(n) = \begin{cases} 1 & \text{, if } n = 1 \\ 2f(\lfloor \frac{n}{2} \rfloor) + n & \text{, otherwise} \end{cases}$$

   Please find out the tightest bound of $f(n)$ with Θ notation. You should provide a proof to get credits. Answers without any proof will NOT get any credits.

   b. (4 *points*) Define a function $T(n)$ as below:

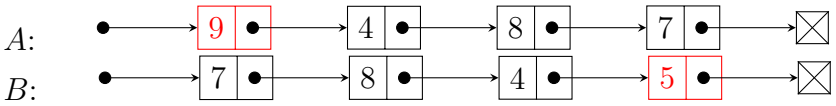
$$T(n) = \begin{cases} 1 & \text{, when } n = 1 \\ 2T(\lfloor \sqrt{n} \rfloor) + \lg n & \text{, otherwise} \end{cases}$$

   Prove that $T(n) = O(\lg n \times \lg \lg n)$. (Hint: Transformation of variables.)

Problem 3. Linked List (25 *points*)

For each problem, if it's not trivial, please briefly explain why your procedure is correct and why the time/space complexity meets the constraints.

1. (5 *points*) You are given the heads of two read-only singly linked lists A and B. Both A and B are loop-free and store positive integers. The integers in A, from head to tail, form a sequence $\langle a \rangle$, and the integers in B, from tail to head, form a sequence $\langle b \rangle$. Guaranteed that $\langle a \rangle$ and $\langle b \rangle$ are of same length n, and there exists exactly one i s.t. $a_i \neq b_i$. Please describe a procedure to find i.



$\langle a \rangle = (9, 4, 8, 7), \langle b \rangle = (5, 4, 8, 7), i = 1, a_i = 9, b_i = 5.$

Your procedure should take the heads of A and B as its input, then output i, while i is the index mentioned above. Both A and B are read-only, you cannot modify them.
In $o(n^2)$ time, $O(1)$ extra space.

Solution.

It's easy to verify that $\forall P \subseteq [1, n], \sum_{j \in P} a_j = \sum_{j \in P} b_j$ iff $i \notin P$.

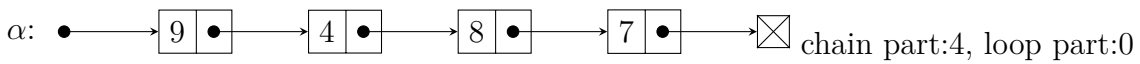
This gives a motivation to start a binary search to find the smallest P that contains i, that is, the set that contains i itself.

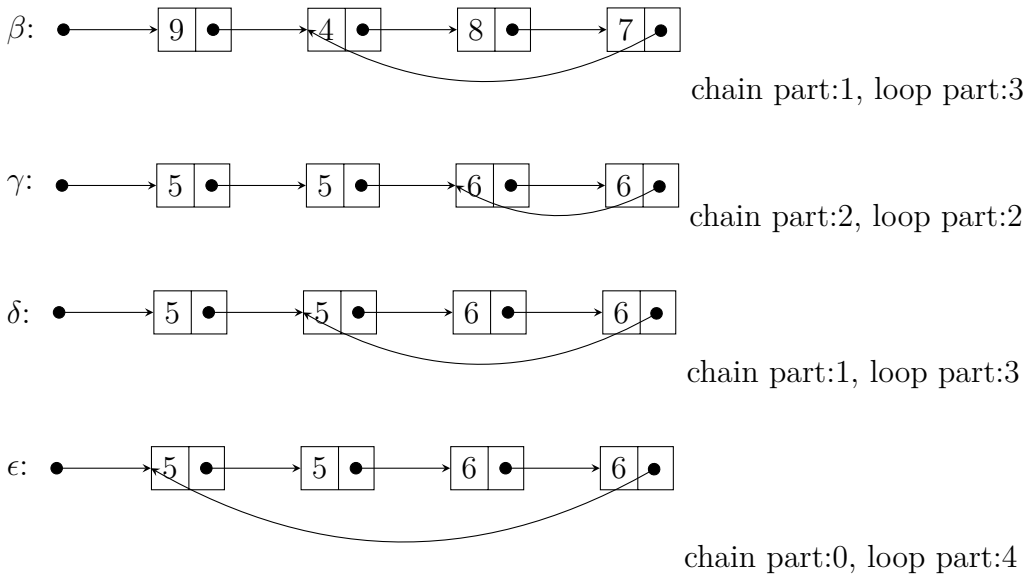
PROCEDURE 4.1($A.head, B.head$)

```
1  leftA ← A.head
2  leftB ← B.head
3  ans ← 1
4  length ← n
5  while length > 1
6      halfLength ← ⌊length/2⌋
7      middleA ← A.head
8      middleB ← B.head
9      sumA, sumB ← 0
10     for i = 1 to halfLength
11         sumA ← sumA + middleA.data
12         sumB ← sumB + middleB.data
13         middleA ← middleA.next
14         middleB ← middleB.next
15     if sumA = sumB
16         leftA ← middleA
17         leftB ← middleB
18         ans ← ans + halfLength
19         length ← length − halfLength
20     else length ← halfLength
21 output ans
```

2. (20 *points*) You are given the heads of two read-only singly linked list A and B. The tail node might not point to *NULL* but point to another node, so they might contain loop. Please follow the guide and design a procedure to determine whether the shapes of the linked lists are identical.

   For example, among the following linked lists, linked list β and linked list δ are identical in shape. You may discover that a linked list contains at most one loop, and if there is one, then the loop must appears at the end. Thus, two linked lists are identical in shape if and only if their "chain parts" and the "loop parts" are of equal sizes respectively, while the data they store are irrelevant to their shapes.





Both A and B are read-only, you cannot modify them.

a. (5 *points*) Please complete the C function below to determine whether the given linked list has a loop. If so, return a pointer to an arbitrary node on the loop. Otherwise, return *NULL*.

In $O(n)$ time, $O(1)$ extra space, while n is the number of nodes of the given linked list. Please note that the only input is the head of the linked list, and you are ignorant of the value of n.

(Hint: Consider two pointers pointing to the head initially and moving at different speed.)

```
typedef struct listnode{
    int data;
    struct listnode *next;
}ListNode;

ListNode* isLoop(ListNode* head){
    ListNode* ptr_slow, ptr_fast=head;
    while(ptr_slow!=ptr_fast&&ptr_fast){
        ptr_slow=ptr_slow->next;
        ptr_fast=ptr_fast->next;
        if(ptr_fast->next)ptr_fast=ptr_fast->next;
    }
    return ptr_fast;
}
```

b. (3 *points*) Please complete the C function below to determine the length of the loop of the given linked list. The length of the loop is defined as the number of nodes on the loop. You are free to use function `isLoop()` defined above.
In $O(n)$ time, $O(1)$ extra space.

Solution.

```
unsigned int loopLength(ListNode* head){
    ListNode* loopNode=isLoop(head);
    if(!loopNode)return 0;
    int l=1;
    ListNode* tmpNode=loopNode->next;
    while(tmpNode!=loopNode){
        l++;
        tmpNode=tmpNode->next;
    }
    return l;
}
```

Please describe a procedure that determine whether A and B are of same shape. Your procedure should takes the heads of A and B as its input. If they are of same shape, output 1, otherwise, output 0. In:

c. (6 *points*) $O(n \lg n)$ time, $O(1)$ extra space.

d. (6 *points*) $O(n)$ time, $O(n)$ extra space.
n is the sum of the numbers of nodes of A and B, and you are ignorant of the value of n. You are free to use functions `loopLength()` and `isLoop()` defined above.

Solution.

For simplicity, let's define two procedure FIND-NODE and FIND-INDEX first.

FIND-NODE($head, index$)
1 $tmpNode \leftarrow head$
2 **while** $index > 0$
3 $\quad index \leftarrow index - 1$
4 $\quad tmpNode \leftarrow tmpNode.next$
5 **return** $tmpNode$

```
FIND-INDEX(head, node)
1  tmpNode ← head
2  index ← 1
3  while tmpNode ≠ node
4        index ← index + 1
5        tmpNode ← tmpNode.next
6  return index
```

FIND-NODE is a procedure that returns the node of the given index. FIND-INDEX is a procedure that returns the index of the given node. It's easy to verify that both procedure comsumes linear time and constant extra space.

Note that two linked lists are identical in shape if and only if their chain parts and loop parts are of same length respectively.In problem (b), we have introcuded a $O(n)$-time and $O(1)$-space procedure `loopLength` to calculate the length of the loop part of a given linked list. Thus, it remains to calculate *chainLength*, the length of the chain part. The followings are two procedures for calculating *chainLength*, and meets the constraints of (c) and (d) respectively.

The procedure CHAIN-LENGTH(C) exploits the fact that if $chainLength > 0$, then *chainLength* is exactly the index of the last node that is not on the loop. Thus, we use a binary search to find such index.

```
chainLengthC(head)
 1  left ← 1
 2  right ← 1
 3  loopNode ← isLoop(head)
 4  right ← FINDINDEX(loopNode)
 5  if loopNode = NULL
 6        return right − 1
 7  while right − left > 1
 8        middle ← ⌊(left + right)/2⌋
 9        if FIND-NODE(head, middle) = FIND-NODE(head, middle + loopLength(head))
10              right ← middle
11        else
12              left ← middle
13  return left
```

```
chainLengthD(head)
 1  loopNode ← isLoop(head)
 2  index ← Find-Index(head, loopNode)
 3  if loopNode = NULL
 4       return index − 1
 5  let chainArray[1 . . index] and loopArray[1 . . loopLength(head)] be new arrays
 6  tmpNode ← head
 7  for i = 1 to index
 8       chainArray[i] ← tmpNode
 9       tmpNode ← tmpNode.next
10  for i = 1 to loopLength(head)
11       loopArray[i] ← tmpNode
12       tmpNode ← tmpNode.next
13  i_chain ← index
14  i_loop ← loopLength(head)
15  while chainArray[i_chain] ≠ loopArray[i_loop]
16       i_chain ← i_chain − 1
17       i_loop ← i_loop − 1
18  return i_chain
```

Programming Part

Problem 4. Notepad -- (20 *points*)

You are required to simulate a very lightweight text editor: *Notepad--*. The following table shows all the effective keys in this editor and their effects. All the other keys are ineffective and **should be ignored** by the editor.

`a-z`	Overwrite the selected segment with the character; if nothing is selected, simply write the character. After that, move the cursor to the end of the inserted character, and exit *selection mode*.
`H`	Move the cursor leftward by one. If the cursor is at the beginning of the file, do nothing.
`L`	Move the cursor rightward by one. If the cursor is at the end of the file, do nothing.
`I`	Move the cursor to the beginning of the file.
`A`	Move the cursor to the end of the file.
`V`	Enter *selection mode*. If currently in *selection mode*, exit *selection mode*.
`D`	If nothing is selected, do nothing. Otherwise, cut the selected segment and overwrite the clipboard with it, and then exit *selection mode*.
`P`	Overwrite the selected segment with the content in the clipboard; if nothing is selected, simply paste the clipboard content. After that, clear the clipboard, move the cursor to the end of the pasted content, and exit *selection mode*.

Note that in *Notepad--*, the cursor is NOT supposed to lie just upon a character; instead, it's always between two characters or at the beginning/end of file. Briefly speaking, it looks like the left figure instead of the right one.

Also note that the selection mode works just like keeping <shift> key pressed in Notepad or Word. More precisely, in *selection mode*, the "selected segment" is the segment between the current cursor and the cursor position when you enter *selection mode*. While not in selection mode, nothing is selected.

Suppose *Arvin* just open an empty file with *Notepad--*, and the clipboard is initially empty. Given the series of keys pressed by *Arvin*, what would the file eventually be?

Input Format

The first line is an integer T, indicating that there are T test cases. For every test case, there

is a string composed of lowercase and uppercase characters, which is the series of keys *Arvin* keys in.

$T \leq 5$.

The length of each string $\leq 10^6$

Output Format

For every test case, output the final content of the file.

Input Constraint

For 30% of the testcases, there won't be any `I`, `A`, `V`, `D` or `P` in the input.

For 70% of the testcases, there won't be any `I` or `A` in the input.

For 100% of the testcases, there are no other constraints.

Sample Input

```
5
tleVHMEOWHHwaVHBBHac
dsahahaVHHHHDggHHHHHPLLLLLgggqaq
notepadVHHHDHHVLHDHHDP
arvinLVHHDHHHHPVLPLDLP
dicxintheIVLLLLmeowAbox
```

Sample Output

```
ac
hahadsagggggqaq
note
inrv
meowinthebox
```

Problem 5. The Crafty TA (20 *points*)

As a crafty TA, *Maoq* would like to give students as few points as possible. The rules of giving points are as below. The student got N points originally, and *Maoq* can adjust his points by removing k arbitrary digits, but this removal shouldn't yield any leading zeros. Given N and k, what's the minimum points the student would get?

Note: A single 0 is counted as leading zero.

Input Format

The input contains 2 integers, N and k. Both N and k will not contain any leading zeros.

$1 \leq N \leq 10^{100000000}$

$1 \leq k < len(N)$

Output Format

An integer, which is the minimum points the student would get.

Input Constraint

For 5% of the testcases, $len(N) \leq 10$, and each digit of N is not zero.

For 30% of the testcases, $len(N) \leq 10^4$, and each digit of N is not zero.

For 50% of the testcases, $len(N) \leq 10^4$.

For 100% of the testcases, there are no other constraints.

Sample Input

```
30400 2
```

Sample Output

```
300
```