NHAY - A Needle in the Haystack

*no tags*

Write a program that finds all occurences of a given pattern in a given input string. This is often referred to as finding a *needle* in a *haystack*.

The program has to detect **all** occurences of the needle in the haystack. It should take the needle and the haystack as input, and output the positions of each occurence, as shown below. The suggested implementation is the KMP algorithm, but this is not a requirement. However, a naive approach will probably exceed the time limit, whereas other algorithms are more complicated... The choice is yours.

Input

The input consists of a number of test cases. Each test case is composed of three lines, containing:

* the length of the needle,
* the needle itself,
* the haystack.

The length of the needle is only limited by the memory available to your program, so do not make any assumptions - instead, read the length and allocate memory as needed. The haystack is **not** limited in size, which implies that your program should not read the whole haystack at once. The KMP algorithm is stream-based, i.e. it processes the haystack character by character, so this is not a problem.

The test cases come one after another, each occupying three lines, with no additional space or line breaks in between.

Output

For each test case your program should output all positions of the needle's occurences within the haystack. If a match is found, the output should contain the position of the first character of the match. Characters in the haystack are numbered starting with zero.

For a given test case, the positions output should be sorted in ascending order, and each of these should be printed in a separate line. For two different test cases, the positions should be separated by an empty line.

Example

**Sample input:**

2

na

banananobano

6

foobar

foo

9

foobarfoo

barfoobarfoobarfoobarfoobarfoo

**Sample output:**

2

4

3

9

15

21

Note the double empty line in the output, which means that no match was found for the second test case.

* Solution:

1. #include<bits/stdc++.h>
2. #define int long long int
3. #define pb push\_back
4. using namespace std;
5. int mod = 1e9+7;
6. int p=31;
7. main()
8. {
9. int n;
10. while(cin>>n)
11. {
12. string s1,s2;
13. cin>>s1>>s2;
14. int hs = 0;
15. int n1 = s1.length(),n2 = s2.length();
17. int x=max(n1,n2);
18. int power[x]={0};
19. power[0]=1;
20. for(int i=1;i<x;i++)
21. power[i]=(power[i-1]\*p) % mod;
23. if(n1>n2)
24. {
25. cout<<"**\n**";
26. continue;
27. }
29. for(int i=0;i<n1;i++)
30. hs=(hs + (s1[i]-'a'+1)\*power[i]) % mod;
32. int has[n2+1] = {0};
33. for(int i=0;i<n2;i++)
34. has[i+1]=(has[i]+(s2[i]-'a'+1)\*power[i]) % mod;
36. vector<int> ans;
37. for(int i=0;i+n1-1<n2;i++)
38. {
39. int curr=(has[i+n1]-has[i]+mod)%mod;
40. if(curr == (hs\*power[i])%mod)
41. ans.pb(i);
42. }
43. sort(ans.begin(),ans.end());
44. if(ans.size()==0)
45. {
46. cout<<"**\n**";
47. }
48. else
49. {
50. for(int i=0;i<ans.size();i++)
51. cout<<ans[i]<<"**\n**";
52. }
53. }
54. }

* Tips:
* Reference -> Cp algorithms

# Rabin-Karp Algorithm for string matching

**Table of Contents**

[Implementation](https://cp-algorithms.com/string/rabin-karp.html#toc-tgt-0)

[Practice Problems](https://cp-algorithms.com/string/rabin-karp.html#toc-tgt-1)

This algorithm is based on the concept of hashing, so if you are not familiar with string hashing, refer to the [string hashing](https://cp-algorithms.com/string/string-hashing.html) article.

This algorithm was authored by Rabin and Karp in 1987.

Problem: Given two strings - a pattern ss and a text tt, determine if the pattern appears in the text and if it does, enumerate all its occurrences in O(|s|+|t|)O(|s|+|t|) time.

Algorithm: Calculate the hash for the pattern ss. Calculate hash values for all the prefixes of the text tt. Now, we can compare a substring of length |s||s| with ss in constant time using the calculated hashes. So, compare each substring of length |s||s| with the pattern. This will take a total of O(|t|)O(|t|) time. Hence the final complexity of the algorithm is O(|t|+|s|)O(|t|+|s|): O(|s|)O(|s|) is required for calculating the hash of the pattern and O(|t|)O(|t|) for comparing each substring of length |s||s| with the pattern.

## Implementation

**vector<int> rabin\_karp(string const& s, string const& t) {**

**const int p = 31;**

**const int m = 1e9 + 9;**

**int S = s.size(), T = t.size();**

**vector<long long> p\_pow(max(S, T));**

**p\_pow[0] = 1;**

**for (int i = 1; i < (int)p\_pow.size(); i++)**

**p\_pow[i] = (p\_pow[i-1] \* p) % m;**

**vector<long long> h(T + 1, 0);**

**for (int i = 0; i < T; i++)**

**h[i+1] = (h[i] + (t[i] - 'a' + 1) \* p\_pow[i]) % m;**

**long long h\_s = 0;**

**for (int i = 0; i < S; i++)**

**h\_s = (h\_s + (s[i] - 'a' + 1) \* p\_pow[i]) % m;**

**vector<int> occurences;**

**for (int i = 0; i + S - 1 < T; i++) {**

**long long cur\_h = (h[i+S] + m - h[i]) % m;**

**if (cur\_h == h\_s \* p\_pow[i] % m)**

**occurences.push\_back(i);**

**}**

**return occurences;**

**}**