

# STRING

- Sequence of characters
- small set
- contiguous integer value 'a' to 'z' and 'A' to 'Z' in both ASCII and UTF-16

c/c++

char:- ASCII  
8 Bit

Java

UTF-16  
16 Bit

Also supports what

```
char x = 'a'  
cout << (int)x; // 97
```

Print the frequencies of character in sorted order

```
string str = "geeksforgeeks";  
int count[26] = {0};  
for (int i=0; i< str.length(); i++)  
    count [str[i] - 'a']++;  
for (int i=0; i< 26; i++){  
    if (count[i] > 0){  
        cout << (char)(i+'a') << " ";  
        cout << count[i];  
    }  
}
```

O/P

e	4
f	1
g	2
k	2
o	1
r	1
s	2.

Strings in C++ (char arrays)

comes in

[g | f | g | \0]

if we give definite size

char str[] = "gfg"

str[6] = "gfg"

[g | f | g | \0 | \0 | \0]

func

strlen → gives length of string

- strcmp (a, b) → dictionary types. compare strings lexicographically res = strcmp(a, b)

- a < b then res < 0
- a = b then res = 0
- a > b then res > 0

### strcmp (a, b)

To copy another string we have to use strcpy

strcpy (str1, "ftg")

str is an address and we can't assign anything directly to an address.

### C++ string

- richer library
- supports ~~opposite~~ operations like +, <, >, ==, !=, <=,
- >= are possible;
- can assign a string later.
- do not have to worry about size
- C++ strings are ~~classes~~ objects of classes.

\* str.length() → length of string

\* str.substring (start, length) →

beginning index

length of string you want

"geekforgeeks" ← (1,3) → geek

"wldsh" ← (2,3) side

compare strings lexicographically res = strcmp(a, b)

str.find ("eek")

return → position

→ returns index of first occurrence.  
→ otherwise string ::npos.

string str = "geekforgeeks"

str.find ("eek"); → 1  
str.find ("kaf"); → 3.

### strcmp

in C++ we have ==, >, <

### Reading string from console

string str;  
cin >> str;  
cout << str;

case 0

were entered - sandeep Jain  
output → sandeep

case 1

were entered → sandeep  
output → sandeep

Reason: whenever were entered an ' ' (space)  
at the ~~or~~ cin operation

stops reading the character.

For this purpose we use getline

string str;  
getline (cin, str)

it will stop reading  
the character after  
you press enter

it can also accept an  
optional character.  
getline (cin, str, 'g')

stop when  
you see 'g'

## Anagram of Each other

→ checking if two strings are permutation of each other  
→ Every character in the first string should be present in the second and the frequency must also be same.

$s_1 = "listen"$ ,  $s_2 = "silent"$ .

O/P → Yes

$s_1 = "aab"$ ,  $s_2 = "bab"$

O/P → NO.

### Naïve Approach O(nlogn)

sort both the string and then compare them.

```
sort(s1.begin(), s2.end())
sort(s2.begin(), s2.end())
return s1 == s2;
```

### Efficient solution

We create a count array and then used characters of string as indexes.

on each occurrence of character in 1st string we increment the count of that character and on each occurrence of same character in second string we decrement the count.

```
bool anagram(string s1, string s2) {
```

```
    // check for lengths
    int count[256] = {0};
    for (int i=0; i < length(); i++) {
        count[s1[i]]++;
        count[s2[i]]--;
    }
    for (int i=0; i < s1.length(); i++) {
        if (count[s1[i]] > 1)
            return 1;
        return -1;
    }
}
```

## Leftmost repeating character

I/P : str = "geeksforgeeks"  
O/P → 0 'g' is the first repeating char.

I/P → str = "ababc"  
O/P = 1 'b' is the first repeating char.

I/P → str = "ababa"  
O/P = 0

### Naïve approach

Run two loops and check if particular character appears in the string after it.

only

### Efficient approach

use characters as indexes and then run another loop to check  $\text{count}[str[i]] > 1 \rightarrow \text{return } i;$

```
const int CHAR = 256;
int getleftmostrep(string str) {
```

```
    int count[CHAR] = {0};
    for (int i=0; i < str.length(); i++) {
        count[str[i]]++;
    }
}
```

```
    i.
```

```
    for (int i=0; i < str.length(); i++) {
        if (count[str[i]] > 1)
            return i;
    }
}
```

## Effective Approach (To solve in one traversal)

We use an array which is initialised as -1 and whenever we see a character we store its first index into that array.

If we see a character whose corresponding value in t-array is not -1, we get to know that this is repeating character to get the leftmost non-repeating true, the minimum of all repeating

```
int leftmost(string str) {
    int f_index[CHAR];
    fill(f_index, f_index + CHAR, -1);
    int res = INT_MAX;
    for (int i = 0; i < str.length(); i++) {
        int fi = f_index[str[i]];
        if (fi == -1)
            f_index[str[i]] = i;
        else
            res = min(res, fi);
    }
}
```

## Effective Approach-2

We traverse from right side.

```
int leftmost(string str) {
    bool visited[CHAR];
    fill(visited, visited + CHAR, false);
    int res = -1;
    for (int i = 0; i < str.length(); i++) {
        if (visited[str[i]] == -1)
            visited[str[i]] = i;
        else
            count[str[i]] = -2;
    }
}
```

## Index of Leftmost Non-Repeating char

I/P → "geekforgeeks";  
O/P → 5  
I/P → aabb  
O/P → -1

### 1st solution

Make an array count and then store frequencies of every character.  
Now just traverse on this array and check if its count is 1 (return i).

### 2nd solution

We will make an array named status.  
if status[str[i]] = -1 (unvisited)  
status[str[i]] = -2 (repeated)  
status[str[i]] = ≥ 0 (element is present only once).

```
int nonrep(string str) {
```

```
    int count[CHAR];
    fill(count, count + CHAR, -1);
    for (int i = 0; i < str.length(); i++) {
        if (count[str[i]] == -1)
            count[str[i]] = i;
        else
            count[str[i]] = -2;
    }
}

for (int i = 0; i < CHAR; i++) {
    res = min(res, count[i]);
}

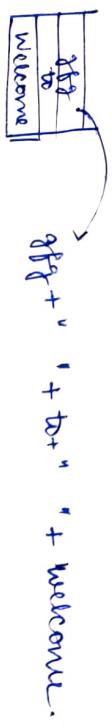
res = max(res, count[i]);
```

## Reverse words in a string

I/P: "Welcome to gfg"  
O/P: "gfg to welcome"

### Naive solution

Push the words inside the stack and then pop the words while popping space.



Auxiliary space  $\rightarrow O(n)$ .

### Effective solution

We reverse individual words and then we reverse the whole string.

I/P  $\rightarrow$  "Welcome to gfg"

"emoclew ot gfg"

gfg to welcome.  $\leftarrow$  required string

void reverseWord (char str[], int n) {

int start = 0; int end = n; int i;

if (str [end] == ' ') {

reverse (start, end - 1, str);

start = end + 1;

} else {

reverse (start, n - 1, str);

however, (0, n - 1, str);

have: 3  
new word  
root word  
word

```
void reverse (char str[], int low, int high) {
    while (low < high) {
        swap (str [low], str [high]);
        low++;
        high--;
    }
}
```

- If we use character array and get its size using size of operator, then we will get one extra as compiler adds '\0' at end, so we can call the function get (str, n-1).

## Overview of pattern searching

I/P - "geeksforgeeks" pattern  $\rightarrow$  "eks"  
O/P = 2, 10

I/P  $\rightarrow$  "AAAAAA"  
O/P  $\rightarrow$  0, 1, 2

I/P  $\rightarrow$  ABCDEFABD      pat.  $\rightarrow$  "ABF"  
O/P  $\rightarrow$  not present

## Pattern searching Algorithm

Naive:  $\Theta((n-m+1)*m)$ .  
 $\Theta(m^2)$  where all characters are distinct

$m \rightarrow$  pattern length  
 $n \rightarrow$  text length  
 $\Theta(m^2)$

no preprocessing

Rabin Karp:  $\Theta((n-m+1)*m)$ .  $\leftarrow$  Preprocess pattern

KMP:  $\Theta(m)$

suffix tree:  $\Theta(m)$

$\leftarrow$  word where we have to find all occurrences of pattern  $\Theta(m^2)$  per recursive call of pattern

## Name pattern searching

We slide the ~~text~~ ~~text~~ pattern over the ~~text~~ and check if it matches with the substring.

```
void patsearching (string text, string pat) {
```

```
    int n = text.length();
    int m = pat.length();
    string curr = text.substring(0, m);
    for (int i = 0; i < n - m; i++) {
```

```
        int n = text.length();
        int m = pat.length();
```

```
        string curr = text.substring(0, m);
```

```
        for (int i = m; i < n; i++) {
```

```
            curr
```

```
        int patsearching (string text, string pat) {
```

```
        int n = text.length();
        int m = pat.length();
```

```
        for (int i = 0; i < n - m; i++) {
```

```
            int j;
```

```
            for (int j = 0; j < m; j++)
```

```
                if (pat[j] != text[i + j])
```

```
                    break;
```

```
                if (j == m)
```

```
                    cout << i << endl;
```

```
}
```

Time comp -  $O((n - m + 1) * m)$

## Improve Name Algorithm → when pattern is distinct

text → ABC A B C D  
pat → AB C D

similar to naive, we try match the pat with the text but try to in naive we move the pattern by 1 every time here as the pattern is distinct we can move the pattern by  $i$  times if  $i$  is the no of char matched.

```
void puncture (string text, string pat) {
```

```
    int n = patsearching.length();
    int m = pat.length();
```

```
    for (int i = 0; i < n - m; i++) {
```

```
        int j = 0;
```

```
        for (j = 0; j < m; j++)
```

```
            if (pat[j] != text[i + j])
```

```
                break;
```

```
            if (j == m)
```

```
                cout << i << endl;
```

```
            if (i == 0)
```

```
                i++;
```

```
        else
```

```
            i = i + j; } mult by i
```

j.

j.

j.

## Rabin-Karp Algorithm

We don't directly compare every window with pattern. We compute the hash of pattern and windows of text. If hash value match we compare the character.

Hash-function → sum of ascii value of characters

Let the ascii values of  $a \rightarrow 1, b \rightarrow 2, c \rightarrow 3$

$$a \rightarrow 1$$

$$d \rightarrow 4$$

$$e \rightarrow 5$$

txt = " abcabcabc"

pat = "abc"

$P + m - 1$  value of pattern  $(1+2+3) = 6$

abababcabc

$$\text{add} \rightarrow 1+2+4 = 7$$

baa  $\rightarrow$  4

dab  $\rightarrow$  7

abc  $\rightarrow$  6 (Match)

cba  $\rightarrow$  7

bab  $\rightarrow$  5

abe  $\rightarrow$  6 (Match)

match match  
next pattern match  
pattern match

match  
Rabin value  
matches but  
char. doesn't

- If hash value of windows matches, we match individual characters.

- We can compute next hash by previous hash - this is also known as rolling hash.

$$t_{i+1} = t_i + \text{txt}[i+m] - \text{txt}[i]$$

- In simple hash, we have problem of spurious hits.

To reduce the chances of matching hash with window which is not same with pattern we use another more effective hash function.

\* It uses the concept of weighted sum.

Let the string be  $a_0a_1a_2\dots a_{n-1}a_n$ . To the hash value for this string would be

$$a_{n-2} \times d^2 + a_{n-1} \times d^1 + a_n \times d^0$$

→ we have to calculate  $\rightarrow$  let it be  $\rightarrow a_{n-2}a_{n-1}a_n$

$d = \text{no. of characters in text}$

assuming that pattern is of size 3.

$$t_{i+1} = d(t_i - \text{txt} \times d^{m-1}) + \text{txt}[i+m]$$

understanding → Let str = "abc" → To move the this formulae.  $\rightarrow$  let str = "bababc" → window 1 stop

ahead.

Also we'll add  $d^m$  and we'll add  $d \times d^0$  we have to subtract multiply  $d \times d^m$   $= d$ .

$$\text{txt} \rightarrow a_{n-1}a_{n-2}\dots a_{t+1}\dots a_2a_1a_0$$

Let current window =  $a_t a_{t-1} a_{t-2}$

$$\text{hash} = \frac{a_t \times d^2 + a_{t-1} \times d^1 + a_{t-2} \times d^0}{d}$$

$$\text{next hash} = \frac{a_{t-1} \times d^2 + a_{t-2} \times d^1 + a_{t-3} \times d^0}{d}$$

- we will store the hash value under modulo  $q$

because hash values  
may be large &  
can produce  
integer overflow

we try to choose  
bigger values because  
set  $q = 13$

hash range  $\rightarrow \underline{(0, 12)}$

- we precompute pattern and first window hash
- we can calculate hash by a simple formula

```
P = 0
for (int i=0; i<m; i++) {
    P = P + pat[i];
}
P = (P * d + pat[m]) % q;

```

stop-2

(Horner's rule)

;

- we also precompute  $(d^{m-1}) \% q \rightarrow$  this value is used  
in calculating next  
window hash.

void RabinKarp (pat, txt, m, n) {

check  
for  
spurious  
hit

int n = 1;

for (int i=0; i<m-1; i++)

t = (t \* d + pat[i]) % q;

int p=0, t=0;

for (int i=0; i<n-m; i++)

p = (p \* d + txt[i]) % q;

t = (t \* d + pat[i]) % q;

t = ((d \* (t - txt[i] \* n) + txt[i+m])) % q;

if (t < 0) t = t + q;

for (int i=0; i< m-N; i++)

BB

## Algorithm

```
void RabinKarp (pat, txt, m, n) {
```

```
    int n = 1;
```

```
    for (int i=0; i<m-1; i++)
```

```
        P = (P * d + pat[i]) % q;
```

```
        t = (t * d + txt[i]) % q;
```

using  
horner's  
rule:

```
for (int i=0; i<=N-m; i++) {
```

```
    if (P == t) {
```

break flag = true;

for (int i=0; i<m; i++)

if (txt[i+i] != pat[i+i]) {

flag = flag false;

break;

if (flag == true) {

print(i);

i < m-m) {

i < m-m) {

calculating  
next  
hash:

```
t = ((d * (t - txt[i] * n) + txt[i+m])) % q;
```

```
i < t < 0) t = t + q;
```

## Constructing longest possible prefix suffix array

$O(p \rightarrow \{0, 0, 1, 2, 3\})$

prefix of "abcd" string  
not in p.  
", "a", "ab", "abc"

suffix of "abcd"

"", "d", "cd", "bcd", "abd"

$O(p \rightarrow \{0, 1, 2, 3\})$

for string ababc

(a) at  $i=0$  prefix  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}$   
suffix  $\rightarrow \underline{\underline{\underline{\underline{\underline{ababc}}}}}$   $\rightarrow \textcircled{0}$

(ab) at  $i=1$  prefix  $\rightarrow \underline{\underline{\underline{\underline{a}}}}$ , "a"  
suffix  $\rightarrow \underline{\underline{\underline{\underline{ab}}}}, \underline{\underline{\underline{a}}}$   $\rightarrow \textcircled{0}$

(aba) at  $i=2$  prefix  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}$   
~~prefix~~  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}$   $\rightarrow \textcircled{1}$

(abaa) at  $i=3$  prefix  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}$   $\rightarrow \textcircled{2}$   
~~prefix~~  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}$   $\rightarrow \textcircled{1}$

ababc at  $i=4$  prefix  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}$   $\rightarrow \textcircled{2}$   
~~prefix~~  $\rightarrow \underline{\underline{\underline{\underline{\underline{a}}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}, \underline{\underline{\underline{\underline{a}}}}$   $\rightarrow \textcircled{1}$

$O(p \rightarrow \{0, 0, 1, 2, 3\})$ .

- for all characters same  $\rightarrow \{0, 1, 2, \dots, n-1\}$ .
- length - n

- for all characters distinct  $\rightarrow \{0, 0, 0, \dots, 0, 0\}$ .

str  $\rightarrow$  "abcabcab"

$O(p \rightarrow \{0, 0, 1, 2, 3\})$

str  $\rightarrow$  "ababab"

$O(p \rightarrow \{0, 0, 1, 2, 3\})$

## Naive Approach

$O(n^3)$

We check for every window whether their prefix and suffix match.

abababab  $n=8$

we start by start index  
ie,  $i=5$  — we know the max. possible length of

ababab

prefix is  $i-1$

we will compare

arr[i] with arr[n - len + i]

ie for  $i=5$  arr[0], arr[1], arr[2], arr[3], arr[4], arr[5]

arr[0], arr[3], arr[4], arr[5], arr[6]

arr[1], arr[4], arr[5], arr[6]

arr[2], arr[3], arr[4], arr[5], arr[6]

int longPrefixAndSuffix (str, n) {

for (int len = n-1; len > 0; len--) {

bool flag = true;

for (int i=0; i < len; i++) {

if (str[i] != str[n - len + i]) {

flag = false;

break;

}

if (flag == true)

return len;

return 0; }

void fillLPS (str, lps[]) {

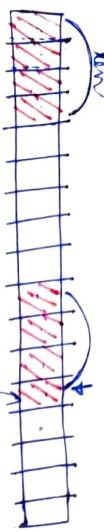
for (int i=0; i<str.length(); i++) {

\*lps[i] = longest proper prefix (str, i+1);

}.

$O(n^3)$

$O(N)$  solution



at  $i=t$

not at  $i=t+1$

\* if  $str[i..n] = str[i..t]$  then  $lps \rightarrow t+1$

$O(N)$  solution

let  $lps[i-1] = len$

then len chars must be matching.

\* if  $str[len] = str[i]$  then  $lps[i] = len + 1$

\* if they do not match

(i)  $len = 0$

then we don't have any common substring  $lps[i] = 0$ .

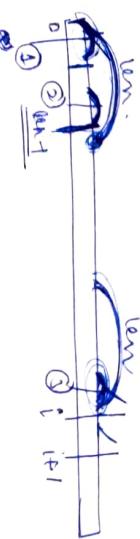
(ii) else

we know that

characters from

0 to  $i-1$  =

char from  $[i-len] \rightarrow i$



we will recursive call  $lps[len-1]$

since (1) = (2)  
and (2) = (3)

so (4) = (3) \* if  $str[lps[i..n]] = str[i]$  then  $lps[i] =$

$lps[len-1]$

we compare  $str[i..n]$  and  $str[i..t]$

\* if  $str[i..n] = str[i..t]$  then  $lps[i] = lps[i..t] + 1$ .

(1)  $\rightarrow$  if  $str[i..n]$  and  $str[i..n]$  match

(2)  $\rightarrow$  if they do not match.

(a)  $len == 0$

$lps[i] = 0$

(b)  $len$

$lps[i] = lps[i..n-1]$

we now compare  $str[i..n]$  and  $str[i..n-1]$  if

they match  $\rightarrow lps[i] = len$  else  
recursively call.

```
void fillLPS (str, lps[]) {
```

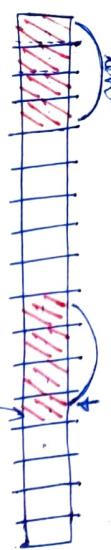
```
for (int i=0; i<str.length(); i++) {
```

```
    lps[i] = longestPrefixMatch (str, i+1);
```

```
}
```

$O(n^3)$

### $O(N)$ solution



at  $i=t$

not at  $i=t+1$

if  $str[i] = str[i]$  then  $lps \rightarrow len+1$

but if  $str[i] \neq str[i]$  then  $lps \rightarrow 0$

~~if  $str[i] \neq str[i]$  then  $lps \rightarrow 0$~~

we compare  $str[i]$  and  $str[i]$

if  $str[i] == str[i]$  then  $lps[i] = lps[i-1] + 1$

but if they are not same

### $O(N)$ solution

let  $lps[i-1] = len$

then  $len$  shall must be matching.

if  $str[i] == str[i]$  then  $\boxed{lps[i] = len+1}$

if they do not match

(i)  $len = 0$

then we don't have any common starting  $lps[i] = 0$

(ii) else

we know that

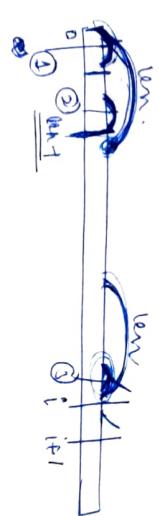
characters from  
0 to  $i-1$  =  
char from  $[i-len]$  to  $i$

we will recursive call  $\underline{lps[i-len]}$

since  $\textcircled{1} = \textcircled{2}$

and  $\textcircled{2} = \textcircled{3}$

so  $\textcircled{1} = \textcircled{3}$  if  $str[lps[i-len]+1] = str[i]$  then  $lps[i] =$



### Basic idea

$\textcircled{1} \rightarrow$  if  $str[i]$  and  $str[i]$  match  $lps[i] = len+1$ .  $len++$

$\textcircled{2} \rightarrow$  if they do not match

(a)  $len == 0$

$lps[i] = 0$

(b)  $len$   
 $len = lps[i-1]$

we now compare  $str[i]$  and  $str[i]$  if

they match  $\rightarrow lps[i] = len$  else  
recursively call.

void fillIPS(string str, ips[]){

int n = str.length(), len = 0;

ips[0] = 0;

int i = 1;

while(i < n){

if(str[i] == str[len]) {

len++;

ips[i] = len;

i++;

else {

if(len == 0) {

ips[i] = 0;

i++;

else {

ips[i] = 0;

i++;

we keep to reducing  
length recursively  
until we face into  
the above two cases.

- \* when characters are distinct  $\Rightarrow O(n)$
- \* when we have all char same  $\Rightarrow O(2n) \rightarrow O(n)$

## KMP String Matching

① KMP is introduced to introduce reduce the running time of algorithm when there are some matches to the pattern.

string agasagaras

m1m2m3

$$\begin{aligned}a_1 &= m_1 \\a_2 &= m_2 \\a_3 &= m_3.\end{aligned}$$

t1t2t3t4t5t6t7t8t9t10t11t12t13

P1P2P3P4P5P6P7P8

We know that  $t_4 = P_1 / t_5 = P_2 / t_6 = P_3$  but  $t_7 \neq P_4$   
as the ips[2] i.e.  $P_3$  is 2 (suppose).

Hence  $P_1P_2 = P_2P_3$ .

also  $t_5t_6 = P_2P_3$

then  $t_5t_6 \approx P_1P_2$

instead of shifting string by 1, we can  
make a shift of  $1 - \text{ips}[i]$ .

- \* we know that - the longest proper prefix that could exist is ips[i].

$\xrightarrow{t_4t_5t_6t_7t_8t_9t_{10}t_{11}t_{12}t_{13}} \quad t_3 \dots t_8 = P_1 \dots P_6$

P1P2P3P4P5P6P7P8

$t_1 \neq P_7$

Instead of shifting by 1 character everytime we just  
check ips[1] because if there is a chance that  
next character will also match with  $P_1$  and so on then  
 $ips[i] \rightarrow ips[i-1]$  i.e. it gives the longest common  
prefixing let the ips of  $P_6 \rightarrow 4$  then

$$\text{max } P_3 \dots P_6 = P_1 \dots P_4$$

$$\text{and } P_3 \dots P_6 = t_5 \dots t_8$$

when  $P_1 \dots P_4 = t_5 \dots t_8$

match is  $t_5$

Name and KMP vacate similar when the all the pattern in the character is distinct

**void KMP (pat, txt)**

```
int n = txt.length();
int m = pat.length();
```

```
int lps[m];
```

```
int i=0, j=0
```

```
while (i < n) {
```

```
    if (pat[i] == txt[i]) { i++; j++; }
```

```
    else if (j == m) { cout << (i - j); j = lps[i - j]; }
```

```
    else { i++; j = lps[i - 1]; }
```

```
}
```

```
int = ababcababaaad
pat = ababaa
```

$i \rightarrow i+1 \rightarrow$

$i = 0, 1, 2, 3, 4$   
 $j = 0, 1, 2, 3$

at  $i=4$  and  $j=4$  they don't match

$j = lps[3] = 2$

we know that the previous will match so we simply move the pattern more.

a b a b c a b a b a d

Now  $i=4$  and  $j=2$  don't match  
 $j = lps[1] = 0$ .

$i=0$

ababcababaaad  
ab ab a

now at  $i=4$  and  $j=0$   
they don't match, we come to 2nd condn.  
 $i++$ .

ababcababaaad

ab ab a

now at  $i=6$  and  $j=0$  they match  
 $i=7, 8, 9, 10$        $i=1, 2, 3, 4$

and  $j=4 == m$   
hence  $\text{cout} << (10 - 4) = 6$

and  $j = lps[3] = 2$

$i=10$  and  $j=3$

ababcababaaad  
ab ab a  
 $i= lps[0] = 0$   
 $j = lps[2] = 1$

they mismatch

ababcababaaad  
ab ab a  
 $i= lps[0] = 0$   
 $j = lps[2] = 1$

they mismatch

ababcababaaad  
ab ab a  
 $i= lps[0] = 0$   
 $j = lps[2] = 1$

mismatch

$i++ \rightarrow 12$  (mismatch)  
 $j = lps[0] = 0$

ababcababaaad  
ab ab a  
 $i= lps[0] = 0$   
 $j = lps[2] = 1$

mismatch

## Time complexity

maximum  $i$  value in the string can be shifted by  $N$  times and the pattern will slide by the step  $n$  times so the complexity will be upper bounded by  $O(2n)$   $\boxed{O(n)}$ .

## Check if strings are rotations

I/P  $\rightarrow$  ABCD & CDAB  
 $\xrightarrow{\text{ABC}} \xrightarrow{\text{BCDA}} \xrightarrow{\text{CDAB}}$

Yes

I/P  $\rightarrow$  ABABA & BAABA  
 $\xrightarrow{\text{ABABA}} \xrightarrow{\text{BAABA}}$

No.

Yes

BAABA

No.

## Naive solution

- rotate string one by one and check if it matches with  $s_2$ . this is  $O(m^2)$  solution
- $O(m) \rightarrow$  to rotate the clockwise by 1.
- $O(n) \rightarrow$  to compare two strings

## Efficient solution

- we can pattern search in a circular manner
- we can concatenate  $s_1$  with itself and then search the pattern in itself

bool anagram(string s1, string s2) {

if (s1.length() != s2.length()) return false;

return ((s1+s1).find(s2) != string::npos);

}

## Anagram Search

We have to check if pattern is present or any of its permutation is present in the text.

I/P: text  $\rightarrow$  "geeksfgeeks"  
 pat  $\rightarrow$  "geeks"

O/P  $\rightarrow$  Yes

I/P: text  $\rightarrow$  "geeksfgeeks"  
 pat  $\rightarrow$  "mack"  $\rightarrow$  true

O/P  $\rightarrow$  No

where present but may not contain contiguous characters

## Naive solution

• run a naive pattern searching algorithm and instead of searching only pattern we search for anagram.

```
bool isPresent(string str, string pat) {
    int n = str.length();
    int m = pat.length();
    for (int i=0; i < m-n; i++) {
        if (isAnagram(str, pat, i))
            return true;
    }
    return false;
}
```

```
bool isPresent(string str, string pat, int i) {
    int count[CHAR] = 0;
    for (int j=i; j < str.length(); j++)
        count[pat[j]]++;
    for (int j=i; j < i+pat.length(); j++)
        count[str[j]]--;
    for (int j=0; j < CHAR; j++)
        if (count[j] != 0) return false;
    return true;
}
```

```
for (int i=0; i < CHAR; i++) {
    if (count[i] != 0) return false;
}
```

## Effective solution

just a modification of naive approach, we make two count arrays ① → pattern  
② → curr-window

initially we compute count array for 1<sup>st</sup> window and then pattern

for in windows we just have to ~~effextt[i-1]~~ → ~~effextt[i-pat]~~

CT [txt[i]] ++;  
CT [txt[i-m]] --;

O(n \* CHAR)

### Effective method

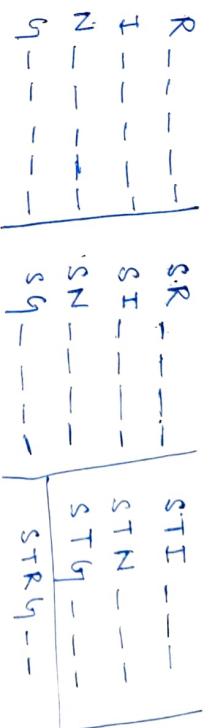
if CT and CP → count arr  
then return for pattern  
base is present (string str, string pat) {  
int CP[CHAR] = {0};  
int CT[CHAR] = {0};  
for (int i=0; i < pat.length(); i++) {  
 CP[tat[i]]++;  
 CP[pat[i]]++;  
}  
for (int i=0; i < pat.length(); i++) {  
 if (CT == CP) {  
 return true;  
 }  
 CT[tat[i]]++;  
 CT[tat[i-pat.length()]]--;  
}

I/P → STRING  
O/P → 598

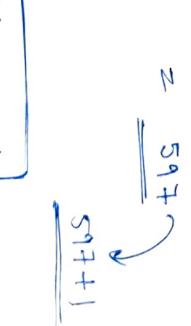
I/P → BAC  
O/P : 3.

A  
B  
C  
→ BAC  
B  
C  
A

- count character which are smaller & which are on right



$$\begin{aligned} & 4 \times 4 + 3 \times 4 \\ & = 120 \times 4 + 120 \times 4 + 6 \times 3 + 1 \\ & = 480 + 96 + 18 + 1 + 2 \\ & = \underline{\underline{597}} \end{aligned}$$



DCBA → 24

A --- | D A -- | D C A - | (DCBA) ✓

B --- | D B --

C ---

$$\begin{aligned} & 3! \times 3 + 2! \times 2 + 1 \times 1 \\ & = 18 + 4 + 1 \Rightarrow (23) \rightarrow \underline{\underline{23+1}} \end{aligned}$$

## Lexicographic Rank of String