

# 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),  
compare strings lexicographically  
res = strcmp(a/b)

if  $a < b$  then res < 0  
if  $a = b$  then res = 0  
if  $a > b$  then res > 0

### strcpy(a/b)

To copy another string we have to use strcpy

strcpy(str, "gfg")

I can't do this  
char str[5];  
str = "gfg";  
cout << str;  
  
str is  
an address.  
and we can't  
assign anything  
directly to an address.

### C++ String

- Richer library
- supports ~~operator~~ 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

"geeksforgeeks" ← (1,3) + eek

"abdef" ← (2,3)  
cde

dictionary types

str.find("eek")

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

string str = "geeksforgeeks"

str.find("eek"); → 1

str.find("ksf"); → 3

### strcmp

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

### Reading string from console

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

#### case 1

user entered → sandeep  
output → sandeep

#### case 2

user entered - sandeep Jain  
output - sandeep

Reason : whenever user entered a ' ' (space)  
at the ~~end~~ cin operator  
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.

$s_1 = \text{"listen"} , s_2 = \text{"silent"}$ .

O/P → Yes

$s_1 = \text{"aab"} , s_2 = \text{"bab"}$

O/P → NO.

### Naive Approach $O(n\log n)$

sort both the string and then compare them.

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

\* we can immediately return false if length is not same

### Efficient Solution

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

on each occurrence of character in 1<sup>st</sup> string we increment the count of that character and on each occurrence of some character in second string we decrement the count.

```
bool anagram(string s1, string s2) {
    // check for length
    int count[CHAR] = {0};
    for (int i=0; i<s1.length(); i++) {
        count[s1[i]]++;
        count[s2[i]]--;
    }
    if (for (int i=0; i<n; i++) if (count[i] != 0) return false)
        return true;
}
```

## Leftmost Repeating character

I/P: str = "geeksforgeeks"

O/P → 0 'g' is the first repeating char.

I/P → str = "abcc"

O/P = 1 'b' is the first repeating

I/P → str = "abcd"

O/P = -1

I/P → str = "abba"

O/P = 0

### Naive Approach

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

$O(n^2)$

### Effective approach

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

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 ( $\text{count}[\text{str}[i]] > 1$ )

return i;

i

return -1;

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 f-array is not -1, we get to know that this is repeating character. To get the leftmost we take the minimum of all repeating.

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

## 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 = str.length() - 1; i >= 0; i++) {
        if (visited[str[i]])
            res = i;
        else
            visited[str[i]] = true;
    }
    return res;
}
```

## Index of leftmost Non-repeating char

I/P → "geeksforgeeks"

O/P → 5

I/P → rabb

O/P → -1

### 1<sup>st</sup> Solution

Make an array count and then store frequencies of every character.

Now just traverse on the array and check if its count is 1 (return i).

### 2<sup>nd</sup> Solution

We will make an array named status.

if  $status[st[i]] = -1$  (unvisited)

$status[st[i]] = -2$  (repeated)

$status[st[i]] \geq 0$  (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,
```

```
        if (count[i] > 0)
```

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

\* Requires one traversal of string.

\* and one traversal of count array (Q56).

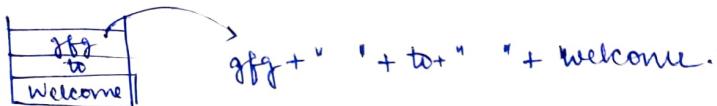
## Reverse Words in a String

I/P: "Welcome to gfg"

O/P: "gfg to Welcome"

### Naive Solution

Push the words inside the stack and then pop the words when while appending 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;

for (int end = 0; end < n; end++) {

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

reverse (start, end - 1, str);

start = end + 1;

}

} reverse (start, n - 1, str);

reverse (0, n - 1, str);

fore.  
reversing  
last  
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, we will get one extra as compiler adds '\n' 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" AAA

O/P  $\rightarrow$  0, 1, 2

I/P  $\rightarrow$  ABCDEFABD pat  $\rightarrow$  "ABF"

O/P  $\rightarrow$  not present

### Pattern searching Algorithm

Naive :-  $O((n-m+1)*m)$ .  $\leftarrow$  no preprocessing

when all characters are  $\rightarrow O(n)$ .  
distinct

$m \rightarrow$  pattern length  
 $n \rightarrow$  txt length  
 $m \leq n$

Robin Karp :-  $O((n-m+1)*m)$ .  $\leftarrow$  preprocesses pattern

KMP :-  $O(n)$

Suffix Tree :-  $O(m)$

used when we have to find all occurrences of pattern  
 $\approx$  thousands of patterns

## Naive Pattern Searching

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

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

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

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

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

```
    int n = txt.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] != txt[i + j])
                break;
        if (j == m)
            cout << i << endl;
```

}

Time comp  $\rightarrow O((n-m+1) \cdot m)$

## Improve Naive Algorithm $\rightarrow$ when pattern is distinct

```
txt → ABC A B C D
pat → A B C D
```

similar to naive, we ~~to~~ match the pat with the text but if ~~a~~ in naive we move the pattern by 1 every time.

Here ~~as~~ as the pattern is distinct we can move the pattern by  $j$  times if  $j$  is the no of char matched.

```
void printchar(string txt, string pat){
```

```
    int n = pat.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] != txt[i + j])
```

```
                break;
```

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

```
            print(i);
```

```
        if (j == 0) } naive
            i++;
        else
```

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

3.  
3.

## 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 characters.

hash-function → sum of ascii value of characters

Let the ascii values of  
 $a \rightarrow 1$        $d \rightarrow 4$   
 $b \rightarrow 2$        $e \rightarrow 5$   
 $c \rightarrow 3$

txt = "abdabcabc"

pat = "abc"

P + hash value of pattern  $(1+2+3) = 6$  !!

abdabcabc      abd  $\rightarrow 1+2+4 = 7$

bda	$\rightarrow 7$	<small>→ no match pattern match</small>
dab	$\rightarrow 7$	
abc	$\rightarrow 6$ (Match)	
bcb	$\rightarrow 7$	
cba	$\rightarrow 6$ (Match) → <small>→ previous txt</small>	
bab	$\rightarrow 5$ Hash value matches but char. doesn't	
abc	$\rightarrow 6$ Match)	

\* If hash value of window 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 hits when window is not same with pattern we use another more effective hash function.

\* It uses the concept of weighted sum.

Let the string be  $a_0a_1a_2a_3\dots a_{t-1}\dots a_{n-2}a_{n-1}$ .

To the hash value for this string would be

abcd

We have to calculate hash for a particular window only.

$$\begin{aligned} &\rightarrow \text{let it be } \rightarrow a_{t-2}x^d + \\ &a_{t-1}x^{d^1} + \\ &a_t x^{d^0}. \end{aligned}$$

Supposing that pattern is of size  $\approx 3$ .

\* We can also use the concept of rolling hash here.

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

Understanding this formulae. → Let str  $\rightarrow$  "abc" → To move the window 1 step ahead.  
 $\rightarrow$   $t \rightarrow$  "babcd" →  $\rightarrow$   $t \rightarrow$  "babcd"

$$\begin{aligned} &\text{also we'll multiply.} \quad \text{and we'll add } d \times d^0 \quad \text{we have to subtract } \\ &d \times (\text{subtracted hash}). \quad = d. \quad \underline{ax^{m-1}}. \end{aligned}$$

$$\text{txt} \rightarrow a_0a_1a_2\dots a_{t-1}\dots a_{n-2}a_{n-1}$$

Let current window =  $a_0a_1a_2\dots a_{t-1}\dots a_{n-2}a_{n-1}$

$$\text{hash} = a_0x^{d^0} + a_1x^{d^1} + a_2x^{d^2} + \dots + a_{t-1}x^{d^{t-1}}$$

$$\text{next hash} = a_{t-1}x^{d^0} + a_{t-2}x^{d^1} + a_{t-3}x^{d^2} + \dots + a_{n-2}x^{d^{n-2}} + a_{n-1}x^{d^{n-1}}$$

- we will store the hash value under modulo  $q$   
 $q \rightarrow$  prime number  
 we try to choose  
 bigger values because  
 let  $q = 13$   
 hash range  $\rightarrow \underline{(0, 12)}$
- because hash value  
 may be large &  
 can produce  
 integer overflow

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

$p = 0$   
 for ( $i=0 \rightarrow m$ )  
 $P = P * d + \text{pat}[i]$   
Step-2      (Horner's Rule)

at $i=0$	$P \leftarrow \text{pat}[0]$
at $i=1$	$P \leftarrow \text{pat}[0]*d + \text{pat}[1]$
at $i=2$	$P \leftarrow \text{pat}[0]*d^2 +$ $\text{pat}[1]*d + \text{pat}[2]$

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

```

void RabinKarp (pat, txt, m, n) {
    int h = 1;
    for (int i = 1; i <= m - 1; i++)
        h = (h * d) % q;
    int p = 0, t = 0;
    for (int i = 0; i < m; i++) {
        p = (p * d + pat[i]) % q;
        t = (t * d + txt[i]) % q;
    }
    for (int i = 0; i < m - N; i++) {
        if (p == t)
            cout << "Pattern found at index " << i;
        p = (p * d + txt[i]) % q;
        t = (t * d + txt[i + N]) % q;
    }
}

```

☆ A

## Algorithm

void RBSearch(const pat, ~~text~~, m, n) {

```

int n=1;
) } N { for (int i=1; i<=m-1; i++)
        n = (n*d)%N;
    }
}

```

int p = 0, t = 0;

```

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

```

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

$\circ\gamma(p == t)\}$

```
bool flag = true;
```

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

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

flag = f  
break;

ii)  $\{H_{20}\} = \{E_{110}\}$

print (i);

calculate  
next  
hash

7

## Constructing longest possible prefix suffix array

I/P → "ababc"

O/P → {0, 0, 1, 2, 0}

I/P → "aaaaa"

O/P → {0, 1, 2, 3}

for string ababc

(a) at  $i=0$       prefix → " "  
suffix → " ", "a"

(ab) at  $i=1$       prefix → " ", "a"  
suffix → "b", "ba", " "

(aba) at  $i=2$       ~~prefix~~ → " ", "a", "ba", "aba"  
P → " ", "a", "ab",

(abab) at  $i=3$       P → "a", "ab", "aba", " "  
S → " ", "b", "ab", "bab", "abab"

ababc at  $i=4$       P → "a", "ab", "aba", "abab", " "  
S → " ", "c", "bc", "abc", "babc",  
"ababc".

O/P → {0, 0, 1, 2, 0}

\* for all characters same → {0, 1, 2, ..., n-1}.

(length=n)

\* for all characters distinct → {0, 0, 0, ..., 0, 0}

str → "abcacabab"

O/P → {0, 0, 1, 0, 1, 2, 3, 0}

str → "ababab"

O/P → {0, 0, 0, 1, 2, 3}

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

suffix of "abcd"      "  
", "d", "cd", "bcd", "abcd"

## Naive Approach

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

For example → let  $i = 5$

we only consider = ababac

we start by last index

i.e.,  $i = 5$  → we know the

max. possible length of  
prefix is  $i-1$

we will compare

arr[i] with arr[n - len + i]  
i.e. for  $i = 5$

a[0], a[5]

a[1], a[2]

a[2], a[3]

; ;

a[0], a[3]

a[1], a[4]

a[2], a[5]

; ;

int longProperPrefix (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; }

len =  $\frac{n}{2}$  //  
(both are  
same)

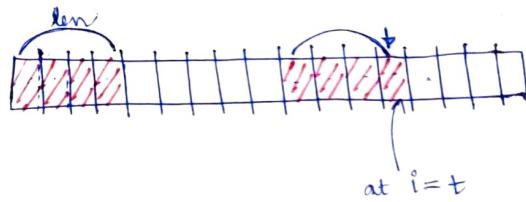
```

void fillLPS (str, lps[]){
    for (int i=0; i<str.length(); i++){
        lps[i] = longProperPrefix (str, i+1);
    }
}

```

$O(n^3)$

### $O(N)$ solution



not at  $i=t+1$

if  $str[len] = str[i]$  then O/P  $\rightarrow len+1$

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

if  $str[len] == str[i]$  then  $LPS[i] = \max(lps[i], lps[i-1]+1)$

but if they are not same

### $O(N)$ 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

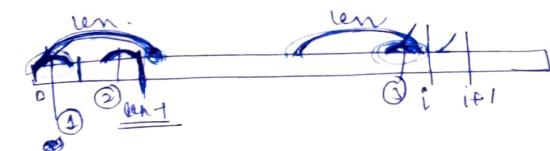
0 to  $i-1$  =  
char from  $[i-len] to i$

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

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

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

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



### Basic Idea

① → if  $str[i]$  and  $str[len]$  match  $lps[i] = len+1$  len++

② → if they do not match.

(a)  $len = 0$

$lps[i] = 0$

(b) else

$len = lps[i-len]$

we now compare  $str[len]$  and  $str[i]$  if they match  $\rightarrow lps[i] = len$  else recursively call  $\underline{lps[i]}$ .

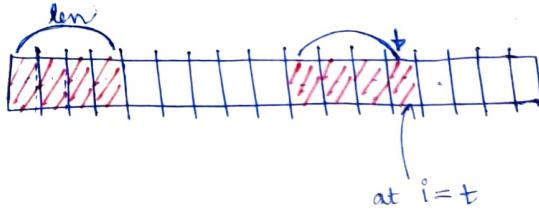
```

void fillLPS (str, lps[]){
    for (int i=0; i<str.length(); i++){
        lps[i] = longProperPrefix (str,i+1);
    }
}

```

$O(n^3)$

### $O(N)$ solution



not at  $i=t+1$

if  $str[len] = str[i]$  then O/P  $\rightarrow len+1$   
but if  $str[len] \neq str[i]$  then O/P  $\rightarrow 0$

~~The contest is that either the length will increase or it will drop to zero~~

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

if  $str[len] == str[i]$  then  $LPS[i] = \max(lps[i], lps[i-1]+1)$

but if they are not same

### $O(N)$ 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

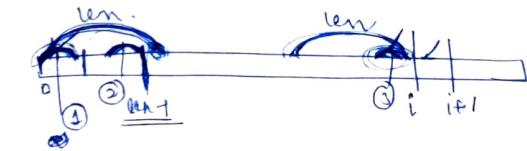
0 to  $i-1$  =  
char from  $[i-len] to i$

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

since  $① = ②$

and  $② = ③$

so  $① = ③$  if  $str[lps[len-1]] = str[i]$  then  $lps[i] = lps[len-1]$



### Basic Idea

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

②  $\rightarrow$  if they do not match

(a)  $len = 0$

$lps[i] = 0$

(b) else

$len = lps[len-1]$

we now compare  $str[len]$  and  $str[i]$  if  
they match  $\rightarrow lps[i] = len$  else  
recursively call.

void fillLPS(string str, lps[]){

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

lps[0] = 0;

int i = 1;

while(i < n){

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

len++;

lps[i] = len;

i++;

}

else {

if(len == 0) {

lps[i] = 0;

i++;

}

else {

len = lps[len];

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

\* when characters are distinct  $\Rightarrow \Theta(n)$

\* when we have all chars same  $\Rightarrow \Theta(n) \rightarrow \Theta(1)$

## KMP String Matching

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

a<sub>1</sub>a<sub>2</sub>a<sub>3</sub>a<sub>4</sub>a<sub>5</sub>a<sub>6</sub>a<sub>7</sub>a<sub>8</sub>  
m<sub>1</sub>m<sub>2</sub>m<sub>3</sub>

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

t<sub>1</sub>t<sub>2</sub>t<sub>3</sub>t<sub>4</sub>t<sub>5</sub>t<sub>6</sub>t<sub>7</sub>t<sub>8</sub>t<sub>9</sub>t<sub>10</sub>t<sub>11</sub>t<sub>12</sub>t<sub>13</sub>  
P<sub>1</sub>P<sub>2</sub>P<sub>3</sub>P<sub>4</sub>P<sub>5</sub>P<sub>6</sub>P<sub>7</sub>P<sub>8</sub>

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

Hence  $P_1P_2 = P_2P_3$ .

also  $t_5t_6 = P_2P_3$

then  $t_5t_6 = P_1P_2$

instead of shifting string by 1, we can make a shift of  $j - lps[j]$ . <sup>(everytime)</sup>

\* We know that the longest proper prefix that could exist is  $lps[j]$ .

t<sub>1</sub>t<sub>2</sub>t<sub>3</sub>t<sub>4</sub>t<sub>5</sub>t<sub>6</sub>t<sub>7</sub>t<sub>8</sub>t<sub>9</sub>t<sub>10</sub>t<sub>11</sub>t<sub>12</sub>t<sub>13</sub>  $\Rightarrow$   $t_3 \dots t_8 = P_1 \dots P_6$   
P<sub>1</sub>P<sub>2</sub>P<sub>3</sub>P<sub>4</sub>P<sub>5</sub>P<sub>6</sub>P<sub>7</sub>P<sub>8</sub>P<sub>9</sub>P<sub>10</sub>  $\neq P_7$

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

means  $P_3 \dots P_6 = P_1 \dots P_4$

and  $P_3 \dots P_6 = t_5 \dots t_8$

so the next char. with which  $P_1$  will match is  $t_5$

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

Naive and KMP execute similarly when the all the pattern in the characters is distinct

```
void KMP (pat, txt) {
    int n = txt.length();
    int m = pat.length();
    int lps[m];
    fillLPS(pat, lps);
    int i=0, j=0
    while (i< n) {
        if (pat[j] == txt[i]) { i++; j++;}
        else {
            if (j == m) {cout << (i - j); j = lps[j-1];}
            else if (i < n && pat[j] == txt[i]) {
                if (j == 0) {i++; j;}
                else {j = lps[j-1];}
            }
        }
    }
}
```

txt = ababcababaad  
pat = ababa

$lps = \{0, 0, 1, 2, 3\}$

$i=0 \rightarrow j++ \rightarrow i=1$

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

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

$j = lps[3] = 2 \quad j=2$

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

ababc<sup>+</sup>cababaad  
ababa

Now  $i=4$  and  $j=2$  don't match

$j=2 \quad lps[1] = 0$

$i=0$

ababc<sup>+</sup>cababaad  
ababa

Now at  $i=4$  and  $j=0$

they don't match, we come to 2nd condn.  
 $i++$ .

abab~~c~~ababaad  
ababa

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

and  $j=4 == M$ .

hence cout << (0 - 4) = 6  
and  $j = lps[3] = 2$

$i=10$  and  $j=3$

ababc<sup>+</sup>cababaad  
ababa

they mismatch  
 $j = lps[2] = 1$

ababc<sup>+</sup>ababaad  
ababa

they mismatch  
 $j = lps[0] = 0$

ababc<sup>+</sup>bababaad  
ababa

$i=10 \rightarrow i=11$  (matched)  
 $j=0 \rightarrow 1$

but

ababc<sup>+</sup>bababaad  
ababa

mismatch  
 $i++ \rightarrow 12$  (loop cond'n false)  
out of loop

$j = lps[0] = 0$

## Time complexity

maximum the i value in the string can be shifted by N times  
and the pattern will slide by the text n times  
so the complexity will be upper bounded by  $O(2n)$

$O(n)$

## Check if Strings are Rotations

I/P  $\rightarrow$  ABCD & CDAB  
Yes

ABCD  $\rightarrow$  B CDA  $\rightarrow$  CDAB

I/P  $\rightarrow$  ABAAA & BAAAA  
Yes

BAAAA

I/P  $\rightarrow$  ABAB & ABBA  
No.

## Naive solution

Rotate string one by one and check if it matches with  $s_2$ . This is  $O(n^2)$  solution

$O(n) \rightarrow$  To rotate the clockwise by 1.

$O(n) \rightarrow$  To compare two strings

## Efficient solution

- ① we can pattern search in a circular manner
- ② we can concatenate  $s_1$  with itself and then search the pattern in itself

```
bool areRotations(string s1, string s2) {
    if (s1.length() != s2.length()) return false;
    return ((s1+s2).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: txt  $\rightarrow$  "geeksforgeeks"  
pat  $\rightarrow$  "frog"  
O/P  $\rightarrow$  Yes

I/P: txt  $\rightarrow$  "geeksforgeeks"  
pat  $\rightarrow$  "seek"  $\rightarrow$  These characters were present but they are not contiguous

O/P  $\rightarrow$  NO

## Naive solution

- If 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; i++) {
        if (areAnagram(str, pat, i))
            return true;
    }
    return false;
}
```

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

## Effective solution

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

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

for i<sup>th</sup> window we just have to  
 $CT[txt[i:i+m]]++;$   
 $CT[txt[i-pat]]--;$   
 $CT[txt[i]]++;$   
 $O(n * CHAR)$

if CT and CP  
match  
then return  
true  
else  
count arr  
for pattern

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

## Lexicographic Rank of String

I/P : BAC

O/P : 3.

A B C  
A C B  
→ B A C  
B C A  
;

I/P → STRING  
O/P → 598

### Effective method

R	S.R	STI	STRIG
I	SI	STN	STRING
N	SN	STG	
G	SG	STRY	

$$4! \times 4 + 3! \times 4 + 3! \times 3 + 1 + 2!$$

$$= 120 \times 4 + 120 \times 4 + 6 \times 3 + 1$$

$$= 480 + 96 + 18 + 1 + 2$$

$$= \underline{597} \rightarrow \underline{597+1}$$

DCBA → 24

A	DA	DCA	(DCBA) ✓
B	DB		
C			

$$3! \times 3 + 2! \times 2 + 1 \times 1$$

$$= 18 + 4 + 1 \rightarrow \underline{23} \rightarrow \underline{23+1}$$

- count character which are smaller & which are on right