

Arrays, Strings & Linked Lists Lecture 6

Sunday, 28 July 2024 3:02 PM

String Matching

$|Pat| = m$ Find all occurrences of pat in txt.

$|Txt| = n$
—x—

Z-algorithm

Txt: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 $n=15$
 a b a c a b a d a b a c a b a

Pat: 0 1 2 3 4 5 6 $m=7$
 a b a c a b a

IPS

Brute force — $\begin{cases} O(m \times n) & \text{Time} \\ O(1) & \text{Space} \end{cases}$

Rabin-karp — $\begin{cases} O(m+n) & \text{Time} \\ O(1) & \text{space} \end{cases}$
Collisions $\triangle!$

KMP algorithm — $\begin{cases} O(m+n) & \text{Time} \\ \underline{O(m)} & \text{space} \end{cases}$

Pat: a b a c a b a

Wt - 7

Z-array

starting at index k,
what is the length of
largest substring which is
a prefix of the original
string.

eg.

	0	1	2	3	4	5	6
	a	b	a	c	a	b	a

Z-array

	0	0	1	0	3	0	1
--	---	---	---	---	---	---	---

LPS

ending at index k
What is the length of
largest prefix which
is also a suffix

How to use z-array to find patterns in text?

Text: a b c a b c a b c a

Pat: a b c a

5, 8, 11
~~5~~ ~~5~~ ~~5~~
 0, 3, 6

[Z-array]

starting at index k,
 what is the length of
 largest substring which is
 a prefix of the original
 string.

New string :- Pat + 'sp char' + Txt

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	a	b	c	a	#	a	b	c	a	b	c	a	b	c	a
<u>z-array</u>	0	0	0	1	0	4	0	0	4	0	0	4	0	0	1

prefix = pattern

Computing Z-array efficiently

String.

Z-array

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
 a a b c a a b c d a a b c a b c a a
 0 1 0 0 4 1 0 0 0 8 1 0 0 6 1 0 0 2 1

$$\text{index} + \text{val} \leq \text{last index}$$

z-array (string s , $|s|=n$)

$z[0] = 0;$

$L \leftarrow 0, R \leftarrow 0$

for ($i: 1 \rightarrow n-1$) {

if ($i > R$) {

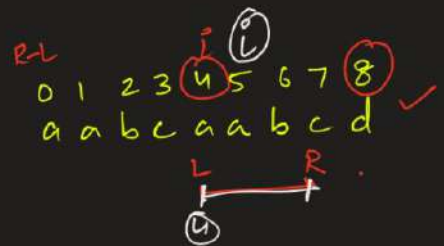
$L = R = i$

while ($R < n$ and $s[R-L] == s[R]$)
 $R++;$

$z[i] = R - L;$

$R--;$

}



$$z[i-L] + i \leq R$$

0 1 2 3 4 5 6 7 8

else {

if ($z[i-L] + i \leq R$)

$z[i] = z[i-L];$ ✓

else {

$L = i;$

while ($R < n$ and $s[R-L] == s[R]$)
 $R++;$

$z[i] = R - L;$ ✓

$R--;$

}

}

}

0 1 2 3 4 5 6 7 8 ✓
a a b c a a b c d ✓
 ↓ ↑ ↓
 4 L R

$i = 5$

$L = 4$

$R = 8$

<https://www.geeksforgeeks.org/problems/search-pattern-z-algorithm-141631/1>

```
vector<int> calc_z(string s) {
    int n = s.length();
    vector<int> z(n, 0);
    int L=0, R=0;
    for(int i=1; i<n; i++) {
        if(i>R) {
            L = R = i;
            while(R<n && s[R-L] == s[R])
                R++;
            z[i] = R-L;
            R--;
        }
        else {
            if(z[i-L]+i <= R)
                z[i] = z[i-L];
            else {
                L = i;
                while(R<n && s[R-L] == s[R])
                    R++;
                z[i] = R-L;
                R--;
            }
        }
    }
    return z;
}
```

$O(n)$

$O(n)$

$R \geq i-1$ ✓

across all loops $O(n)$

```
vector<int> search(string pat, string txt) {
    string ns = pat + "#" + txt;
    vector<int> z = calc_z(ns);
    vector<int> ans;
    for(int i=0; i<z.size(); i++)
        if(z[i] == pat.length())
            ans.push_back(i-pat.length());
    return ans;
}
```

$n+m+1$

$O(n+m)$

$T = O(n+m)$

$S = O(n+m)$

==x==