

Строки

$s[0..n-1]$ $n=|s|$ $s[i] \in \Sigma$ - алфавит

$|s| = 2, 4, 26, 256, \dots$ неоп.

AGTC

Префикс $s = s[0..r]$ $r \leq n-1$

Собственный пр. - $r=n-1$

Сuffix $s = s[l..n-1]$ $l \geq 0$

- " - сфф. - $l=0$

Подстрока $s = s[l..r]$

Поиск подстроки в строке (KMP): если мы знаем, какую часть строки мы ищем, то можем найти её быстрее.

needle

haystack



Наивное решение:

```
for i = 0..|s|-|t|
    ok = True
    for j = 0..|t|-1
        if s[i+j] != t[j]:
            ok = False
            break
    if ok:
        return true
```

$$O(|t||s|) \rightarrow O(|t|+|s|)$$

C++: `strstr find` → быстрый поиск

Python: `in find`

1. З-функция

$lcp[s][i]$ - largest common prefix $s[0..i]$ и $s[i..n-1]$



$$lcp[s][i] = \begin{cases} 0, & s[0] \neq s[i] \\ lcp[s][i-1], & s[0] = s[i] \end{cases}$$

— посчитаем за $O(|s|^2)$

$$z[i] = lcp[s][i]$$

$z[i] = \max k \geq 0: s[0..k] = s[i..i+k]$

- Как с помощью z-функции найти $s \neq t$ t - substring, var. $0 \leq s \leq t$
- 1) посчитаем $z_w[0..|t|-1]$
- 2) for $i = |s|..|t|-1$ if $z_w[i] = |s|$: выходим с позиции i



$$z[i] \geq |s|, \Rightarrow s$$

- Как посчитать z-функцию? $z[0] = 0$
- for $i = 1..|s|-1$: $k = 0$
- while $s[i+k] = s[k]$: $k++$
- $z[i] = k$

$$O(|s|)$$



$$y \& \ell < i \leq z \& \ell = r$$

$$z[i] = \min(z[i-\ell], r-i)$$

$$r = \ell + z[\ell]$$

```
z[0] = 0, l = r = 0
for i = 1..|s|-1:
    k = max(min(z[i-l], r-i), 0)
    while s[i+k] = s[k]:
        k++
    z[i] = k
    if i + z[i] > r:
        l = i, r = i + z[i]
```

Увеличим: ℓ - такое значение i : $\ell + z[\ell] \rightarrow \max$

(*) $\text{len} \Rightarrow r$ увеличивается (увеличится), т.е. $\text{len} \cdot (+)$

(*) $r = i + z[i] = i + k$

т.е. если сделаем r больше (*) применим (*)

r всегда $\in [0..|s|-1] \Rightarrow \sum \text{увелич. } r \leq |s| \cdot |s| \Rightarrow \sum k \leq |s| \cdot |s|$

Time = $O(|s|)$ Mem = $O(|s|)$

Поиск s в t : Time = $O(|s|+|t|)$ Mem = $O(|s|+|t|)$

2. Префикс-функция $\text{нулевые} \leq 1 \quad s[1..|s|]$

s - бордер s , если t -префикс s и t -сuffix s собственный бордер - ts

$\pi[s][i] = \max$ собствен. бордер строки $s[1..i]$

$\pi[s][i] = \max k < i: s[1..k] = s[i-k+1..i]$

Поиск подстроки в строке - алгоритм KMP (Knuth, Morris, Pratt)



Поиск π -ф-ции s

```
 $\pi[s][1] = 0; k = 0$ 
for  $i = 2..|s|$ :
     $k = \pi[s][i-1]$ 
    // переходим бордер строки  $s[1..i-1]$ 
    // и пытаемся добавить  $s[i]$ 
    while  $k > 0$  and  $s[i] \neq s[k+1]$ :
         $k = \pi[s][k]$ 
    if  $s[i] = s[k+1]$ :
         $k++$ 
     $\pi[s][i] = k$ 
     $k$  увелич.  $(+)$   $\Rightarrow k$ 
     $k$  уменьш.  $(-)$   $\Rightarrow k$ 
     $\forall i \neq (k) \Rightarrow \sum k \leq |s|$ 
     $\Rightarrow \sum k \leq |s|$ 
    Time =  $O(|s|)$  Mem =  $O(|s|)$ 
```



Бордер строки $s[1..i]$ - макс. бордер строки $s[1..i]$ + символ $s[i]$



Бордер s : $k \rightarrow \pi[s][k] \rightarrow \pi[s][k]$ Так будет работать все бордер.



3. Полиномиальный хеш

$h_{m,x}(s) = (s[0] \cdot x^{|s|-1} + s[1] \cdot x^{|s|-2} + \dots + s[|s|-1] \cdot x^0) \bmod m$

Реш. не браться по модулю и $x \geq |s|$, то $s = ts \Rightarrow h(s) = h(t)$

Реш. $h_{m,x}$ sat $\Rightarrow h(s) = h(t)$ sff, но $h(s) = h(t)$ - конусный хеш $m \sim 10^8$

sff $\Pr \{h_{m,x}(s) = h_{m,x}(t)\} \sim \frac{1}{m}$

$O(1)$ get_hash(L, r): // [L, r]

return $(ph[s][r] - ph[s][L-1] \cdot x^{r-L+1}) \bmod m$

$ph[s][0] = 0$ $deg[r-L] \quad ph[i] = h_{m,x}(s[0..i])$

for $i = 0..|t|-1$ $ph[s][i+1] = (ph[s][i] \cdot x + t[i]) \bmod m$

deg[0] = 1 for $i = 0..|t|-1$: $deg[i+1] = (deg[i] + x) \bmod m$ $deg[i] = x^i \bmod m$

$s[t] \rightarrow$ считаем $ph[s][0..|t|-1]$ get_hash

$h_{m,x}(s)$ - посчитаем один раз

for $i = 0..|t|-1$ if $hs = \text{get_hash}(i, i+s)$: выходим, выходим

Time = $O(|t|+|s|)$ Mem = $O(|t|) \rightarrow O(1)$

$m \in \mathbb{P} \sim 10^8$ перемен. b - int32 x - const. mod 100

Хеш-таблица

UnsortedMap

массив $xs[0..m]$

$a[x] = y$ add

10^8 $a[x] = ?$ get

x in a ?

$xs[0..m]$

x - строка

$(a[x] = y) \text{ assign}(x, y)$

$a[h(x)] \leftarrow y$

массив $h(x) \in [0..|t|-1]$

1) a-массив смислов h - таблица с зафиксированной адресацией

$a[i] = \{(x, y) \dots \mid h(x) = i\}$ assign(x, y)

assign(x, y): $i = h(x)$ (x, y)

if $(x, y) \in a[i]$ для какого-то y

то заменим $elem(x, y)$

else: $a[i].append((x, y))$

get(x): // find(x)

$i = h(x)$

if $(x, y) \in a[i]$ для x - то y

return y

else return None

2) x - стр. адрес.