

$s == t?$

$s[l_1:r_1] == t[l_2:r_2]$

$137 == 137$

а в а с - представим как число со см. 256

$a = 97 \quad \text{ord}('a') = 97$
 $b = 98$
 $c = 99$

$s = abac \quad h(s) = 97 \cdot 256^3 + 98 \cdot 256^2 + 97 \cdot 256 + 99 < 256^4$

$s = t \Leftrightarrow h(s) = h(t)$
 $h(s) < 256^l, l = |s|$
 Такое $h(s)$ будет числом длины $\Theta(l)$, т.е. будет сравн. за $\Theta(l)$.

Рассм. функцию ф-ии:

$h_m(s) = (97 \cdot 256^3 + \dots + 99) \% m$ Большое $\sim 10^8$
m - простое число Пример

$h_{m,x}(s) = (97 \cdot x^3 + 98 \cdot x^2 + \dots + 99) \% m$ - полиномиальный хеш x=256 m=2

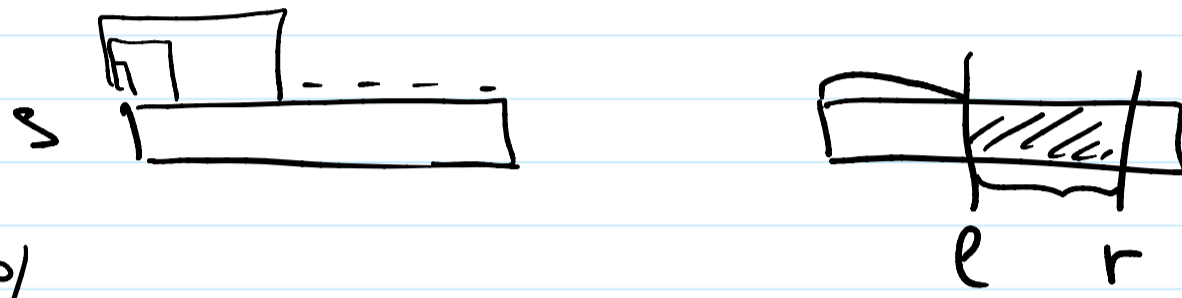
$s = ab \quad h_m(s) = (97 \cdot 256 + 98) \% 2 = 0$

$t = cba \quad h_{m,x}(t) = (99 \cdot 256^2 + 98 \cdot 256 + 100) \% 2 = 0$ } коллизия
s ≠ t
h(s) = h(t)

$s == t \Rightarrow h_{m,x}(s) = h_{m,x}(t)$

$h_{m,x}(s) = h_{m,x}(t) \Rightarrow$ с большой вероятн. $s == t$

Хешы на префиксах



$hp[0] = 0$

$hp[i] = h(s[0..i]) = (hp[i-1] \cdot x + s[i]) \% m$

$hp[l] = (s[0] \cdot x^{l-1} + s[1] \cdot x^{l-2} + \dots + s[l-1] \cdot x^0) \% m$

$hp[r] = (s[0] \cdot x^{r-1} + s[1] \cdot x^{r-2} + \dots + s[l-1] \cdot x^{r-l} + s[l] \cdot x^{r-l-1} + \dots + s[r-1] \cdot x^0) \% m$

$(hp[r] - hp[l] \cdot x^{r-l}) \% m = h(s[l..r])$

В начале посчитаем $x^0, x^1, \dots, x^n \pmod m$

```

s      n = len(s)
hp = [0] * (n+1); deg = [0] * (n+1); deg[0] = 1
for i in range(n):
    hp[i+1] = (hp[i] * x + ord(s[i])) % m
    deg[i+1] = (deg[i] * x) % m
    
```

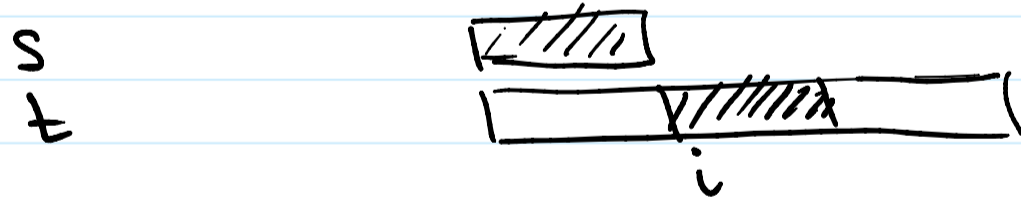
```

def getHash(l, r): # h(s[l..r])
    return (hp[r] - hp[l] * deg[r-l]) % m
    
```

$\leq m^2$ m ~ 10⁸
 $\leq 10^{36} - 2^{120}$ C++: int128_t
 $\% m_1$ m₁ = 10⁹ + 7 (h₁, h₂) = (h₁, h₂)
 $\% m_2$ m₂ = 10⁹ + 9

Итого: сравним подстроки за $O(1)$.

Поиск подстроки в строке



- считаем хеш всей s - h_s $O(|s|)$
- считаем хешы на префиксах t $O(|t|)$
- перебираем i и проверяем, что $h_s == \text{getHash}_t(i, i+|s|)$ $O(|t| \cdot 1)$

Итого $O(|s| + |t|)$ времени
 $O(|t|)$ памяти $\rightarrow O(1)$

Недостатки по сравнению с z-функцией и КМП (П-ф-ией):

- Большая конст. времени работы
 - вероятн. решение
- $|t|$ сравнений
 $\text{Pr}[\text{ошибка}] \leq \frac{|t|}{m} = \frac{n}{m}$ n ~ 10⁶
m ~ 10⁸
 $\text{Pr} \sim 10^{-12}$

Если хешы совпали, можно жестко проверить равенство.

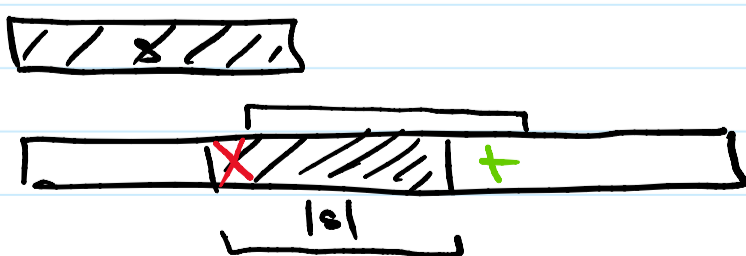
$s \neq t$

$h_{m,x}(s)$ - случайное число 0..m-1
 $h_{m,x}(t)$ - случайное число 0..m-1

$\text{Pr}[\text{коллиз.}] = \frac{1}{m}$ (при одном сравн.)
 $m \sim 10^8$

Алгоритм Равина-Карпа

Поиск подстроки в строке хешами с $O(1)$ памяти



$\text{deg}_s = x^{|s|-1}$
 храним хеш h с текущего окошка $t[i..i+|s|)$

$i \rightarrow i+1$:
 $h_c = (h_c - t[i] \cdot \text{deg}_s + t[i+|s|]) \% m$

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