String matching

$$P[j] = T[i+j] \quad \forall j, \quad 0 \le j \le m-1$$

i

0101111100111 1 — $T \rightarrow n$

$$\sum_{i=0}^{m-1} 2^i \, P[m-1-i]$$

—— $P \rightarrow m$

$\underline{1001}_{\,i_0 i_1 \cdots P_{m-1}} \longrightarrow 9$

for $i \rightarrow 0$ to $n-m$

```
for j -> 0 to m
    if (T[i+j] != P[j]) break;
if (j == m) Print (there is a match at i)
```

J

i $\quad i+j$

j

$\boxed{101011}\,0\,11011$ $\quad 1+2+8+32=43 \quad 86-64$

$2+4+16=22$

101010

$P \Rightarrow 10^6$ $\quad$ 0 1 2 3 4 5

long no. $\quad \underline{101011}$ $\quad 1+2+8+32 = \underline{43}$

$N = P[0]\ ;\ y = T[0] \quad x = 2$ $\qquad N = \sum_{i=0}^{m-1} 2^i P[m-1-i]$

for $j \to 0$ to m

$\quad x = (x \times 2) \% 1$

$\quad N = (N \times 2 + P[j]) \% 1$

$\quad y = (y \times 2 + T[j]) \% P$

if $(N == y)$ ✓ - -

for $(i \to 1$ to $n-m)$

$\quad y = (y \times 2 + T[i+m-1] - X \cdot T[i-1]) \% P$

if $(N == y)$ print (there is a match at i) ✓

$x = y \Rightarrow h(x) = h(y)$

$\cancel{x = y}$ $\quad N \quad y$

$h(N) = N \bmod P \quad h(y) = y \bmod P$

$P(x) \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots a_n x^n$

$a_n x^2 + a_{n-1} x + a_{n-2}$

$P(X_0)$

$t = x \quad S = a_0$

$S = a_n$

for $i \to \cancel{1}$ $n+1$

$S = S + A[i] \times t$

$t = t \times \cancel{x}_0$

$j \to n-1$ to 0

$S = S \times X_0 + A[i]$

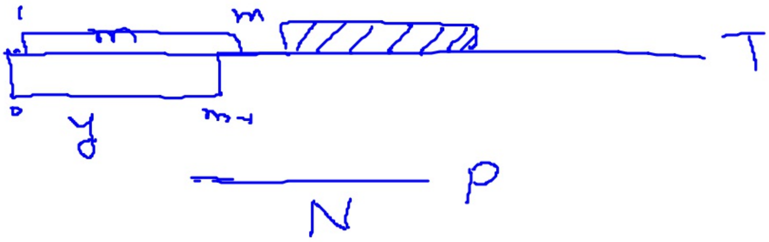
$N = \sum_{i=0}^{m-1} 2^i \cdot P[m-1-i]$

J → 0 to m

if (T[i+j] != P[j]) break;

if (J == m) Point (there is a match at i)

Robin-Karp Algorithm



$n = 10^5$

$m = 10^4$

P matches T at i

$P[j] = T[i+j]$

$j = 0, 1, \ldots, m-1$

$\boxed{1011}1011$

$1011 \to 11$

$0111 \to 7$

$$N = \sum_{i=0}^{m-1} 2^i \cdot P[m-1-i]$$

2^{10000}

$10111 \to 23$

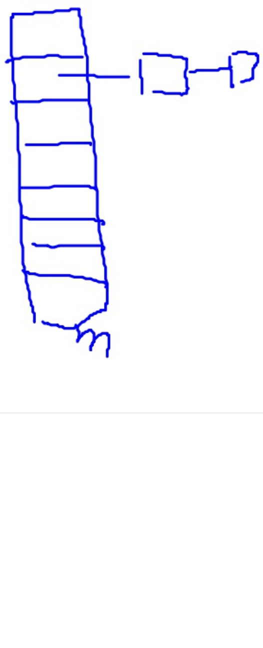
$0111 \to 23-16$

$$h : D \longrightarrow \{0, \ldots m-1\}$$

$$h(x) = x \bmod p$$

Hash table.

$H(i)$



T

rolling hash.

$P[k][m]$

$h(P_1)$ mod

P_1

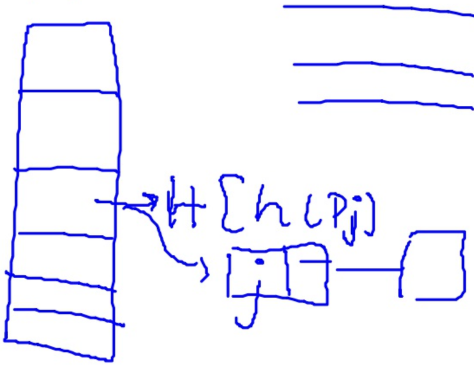
P_2

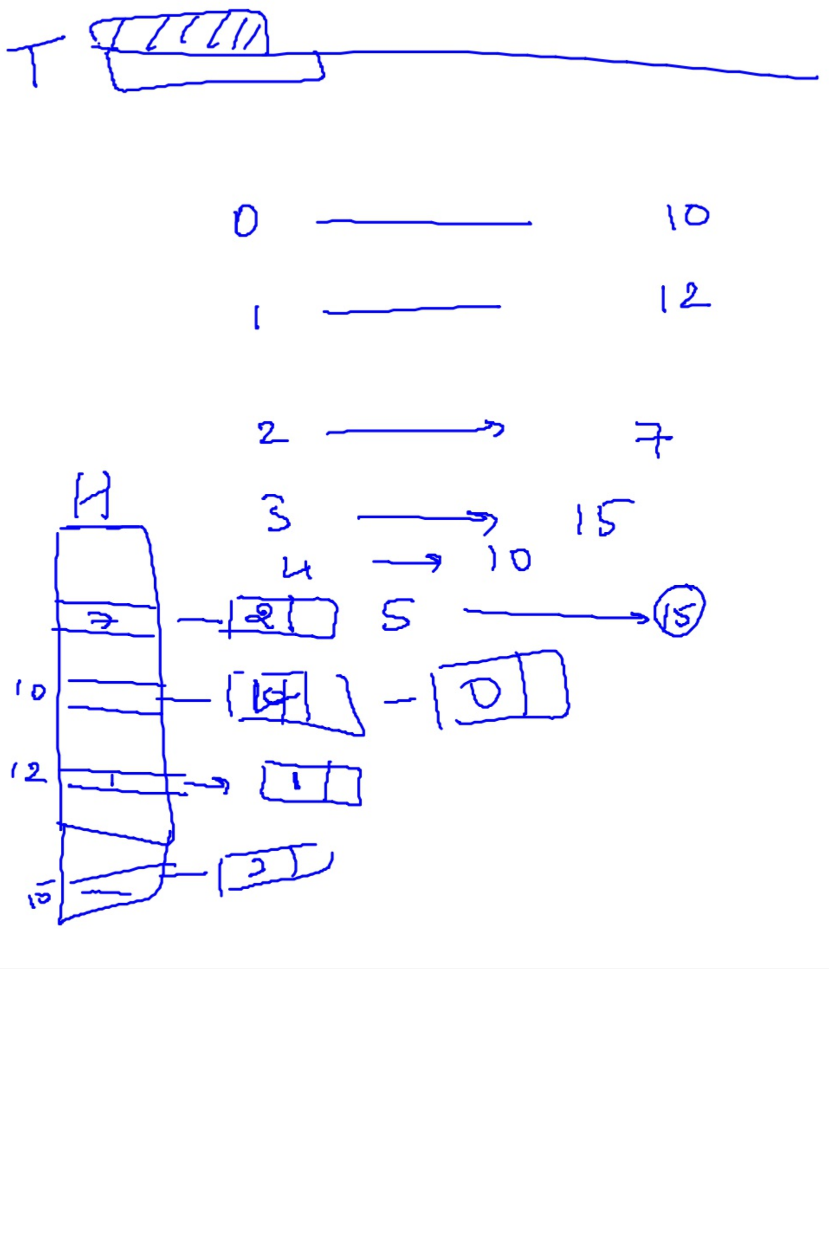
P_k

P_i's $\to m$

$H[h(P_1)]$

H





10 days