Rabin Karp

Ex: Assume you have a string:

'3'	'1'	'4'	'1'	'5'

→ Convert it to number i.e. decimal number

→ m = 5 : Length of string

→ '3' - '0' = 3

'3' → Ascii of '3'

'0' → Ascii of '0'

Integer value 3

→ p = 0

for i = 1 to 5 (i.e. m)

{

p = 10 * p + (P[i] - '0') ; } O(m)

}

Trace Initially p = 0

i = 1 p = 10 * 0 + ('3' - '0')

= 3

i = 2 p = 10 * 3 + ('1' - '0')

= 30 + 1

= 31

i = 3 p = 10 * 31 + ('4' - '0')

= 314

i = 4 p = 10 * 314 + ('1' - '0')

= 3141

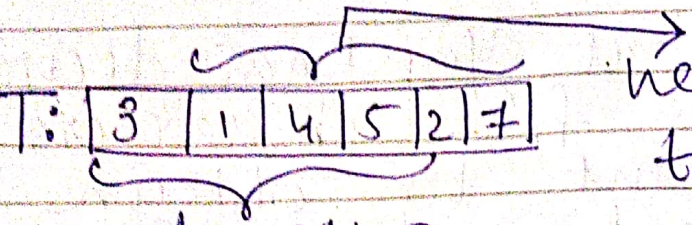
i = 5 p = 10 * 3141 + ('5' - '0')

= 31415

d = Radix = 10

Ex For shift $s=0$, Value is t_0
→ We want Value corresponding to shift $s=1$ i.e. t_1 using t_0.

for Example



we want $t_1 = 14527$ to be Computed using t_0's Value.

$$t_1 = (31452 - 3*10^4)*10 + 7$$
$$= (t_0 - T[1]*10^4)*10 + T[6]$$
$$\therefore t_{s+1} = (t_s - T[s+1]*10^{5-1})*10 + T[s+5+1]$$

Here 5 = Length of pattern
= m

$$\therefore t_{s+1} = (t_s - T[s+1]*10^{m-1})*10 + T[s+m+1]$$
$$= d(t_s - T[s+1]*10^{m-1}) + T[s+m+1]$$

→ Similarly, we will be able to get t_2, t_3 and so on.

Note: Whenever I write $T[i]$, you need to understand it as $T[i] = 'i'$ i.e. Integer Value associated with it.

Algo:

Rabin-Karp-Matcher (T, P, d, q)

{

I/P: T : Text Array T[1..n]

P : Pattern P[1..m]

d : Radix 10 (For decimal)

q : Some Number used for Modulo

Steps: ① n = T.length

m = P.length

$h = d^{m-1} \bmod q$

p = 0

$t_0 = 0$

② for i = 1 to m

Preprocessing {

$p = (dp + P[i]) \bmod q$;

$t_0 = (dt_0 + T[i]) \bmod q$;

}

③ for s = 0 to n-m

{

if (p == t_s)

{

if P[1..m] == T[S+1 ... S+m]

Print "Pattern Occurs with Shift s"

}

if (s < n-m)

{$t_{s+1} = (d(t_s - T[s+1]h) +$

$T[s+m+1]) \bmod q$

}

}

String Matching

}