Analysis of Algorithms

[5CS4-05/5IT4-05]

Unit 3. Pattern Matching

Rabin Karp Algorithm

Ramakant Soni

Assistant Professor, Computer Science Department,

B K Birla Institute of Engineering & Technology, Pilani

Rajasthan

Q1 Write Rabin Karp algorithm for string matching, Represent the same for the given input

String : 9 8 7 6 5 4 3 2 1

Pattern : 5 4 3

Ans:- This algorithm uses notion of equivalence of two numbers modulo a third number.

Pattern and same length characters from string are converted to decimal number of length equal to Pattern.

For Pattern, p is computed using Horner's Rule, it takes time $\Theta(m)$

$$p = P[m] + 10\,(P[m-1] + 10\,(P[m-2] + \dots + 10\,(P[2] + 10\,P[1]))\dots)$$

$$\text{or } p = \sum_{i=1}^{m} P[i] * 10^{m-i}$$

Similarly t_0 will be computed, to compute next values, we use

$$t_{s+1} = 10\,(t_s - 10^{m-1} \cdot T[s+1]) + T[s+m+1]$$

Algorithm for Rabin Karp string matching :-

```
Rabin-Karp-Matcher ( T, P, d, q )
n = T.length
m = P.length
h = d^(m-1) mod q
p = 0
t0 = 0
for i = 1 to m
      p = (dp + P[i]) mod q
      t0 = (dt0 + T[i]) mod q
for s = 0 to n-m
    if p == ts
        if P[1..m] == T[s+1 .. s+m]
            print "Pattern occurs at shift" s
    if s < n-m
        ts+1 = (d(ts - T[s+1]h) + T[s+m+1]) mod q
```

Assuming a d-ary alphabet $\{0, 1, \ldots, d-1\}$, $h \equiv d^{m-1} \pmod q$

Now computing $p = (5 \times 10^2 + 4 \times 10^1 + 3 \times 10^0) \bmod q$

Taking q a prime no., lets $q = 5$

$p = 543 \bmod 5 = 3$

for ~~s~~ $i = 0$

$t_0 = (9 \times 10^2 + 8 \times 10 + 7 \times 10^0 = 987) \bmod q \Rightarrow 987 \bmod 5 = 2$

for $i = 1$ $\quad 10(t_0 - T[1] \times 10^{3-1}) + T[0+1+3]$

$10(987 - T[1] \times 100) + T[4]$

$10(987 - 900) + 6$

$870 + 6$

$t_1 = (876) \bmod 5 = 1$

Similarly $t_2 = 765 \bmod 5 = 0$

$t_3 = 654 \bmod 5 = 4$

$t_4 = 543 \bmod 5 = 3$

$t_5 = 432 \bmod 5 = 2$

$t_6 = 321 \bmod 5 = 1$

for $s=0$ $\quad$ ~~t~~ $p == t_0$ is false because $3 \neq 2$

$s=1$ $\quad$ $p == t_1$ is false because $3 \neq 1$

$s=2$ $\quad$ $p == t_2$ is false because $3 \neq 0$

$s=3$ $\quad$ $p == t_3$ is false because $3 \neq 4$

$s=4$ $\quad$ $p == t_4$ is True because $3 == 3$

$\therefore$ Now we compare $P[1 \ldots m]$ with $T[s+1 \ldots s+m]$
so $5, 8, 3$ compared to $5, 4, 3$
Result is TRUE $\therefore$ Pattern occurs at shift $= 4$

$s=5$ $\quad$ $p == t_5$ is false because $3 \neq 2$

and $s=6$ $\quad$ $p == t_6$ is false because $3 \neq 1$

so, finally we here got the valid shift = 4, where pattern occurs in the given shift.