

Design and Analysis of Algorithms

Lecture-43

Dharmendra Kumar (Associate Professor)
Department of Computer Science and Engineering
United College of Engineering and Research,
Prayagraj

Knuth-Morris-Pratt(KMP) algorithm

Prefix function for a pattern

Given a pattern $P[1..m]$, the *prefix function* for the pattern P is the function $\pi : \{1, 2, 3, \dots, m\} \rightarrow \{0, 1, 2, \dots, m-1\}$ such that

$$\pi(q) = \max\{k \mid k < q \text{ and } P_k \sqsupseteq P_q\}$$

$\pi(q)$ is the length of the longest prefix of P that is a proper suffix of P_q .

Knuth-Morris-Pratt(KMP) algorithm

Example: Compute the prefix function of the pattern

P = ababababca

Solution:

[illegible]

Knuth-Morris-Pratt(KMP) algorithm

KMP-MATCHER(T, P)

```
1   $n = T.length$ 
2   $m = P.length$ 
3   $\pi = \text{COMPUTE-PREFIX-FUNCTION}(P)$ 
4   $q = 0$  // number of characters matched
5  for  $i = 1$  to  $n$  // scan the text from left to right
6      while  $q > 0$  and  $P[q + 1] \neq T[i]$ 
7           $q = \pi[q]$  // next character does not match
8      if  $P[q + 1] == T[i]$ 
9           $q = q + 1$  // next character matches
10     if  $q == m$  // is all of  $P$  matched?
11         print "Pattern occurs with shift"  $i - m$ 
12          $q = \pi[q]$  // look for the next match
```

Knuth-Morris-Pratt(KMP) algorithm

COMPUTE-PREFIX-FUNCTION(P)

```
1   $m = P.length$ 
2  let  $\pi[1..m]$  be a new array
3   $\pi[1] = 0$ 
4   $k = 0$ 
5  for  $q = 2$  to  $m$ 
6      while  $k > 0$  and  $P[k + 1] \neq P[q]$ 
7           $k = \pi[k]$ 
8      if  $P[k + 1] == P[q]$ 
9           $k = k + 1$ 
10      $\pi[q] = k$ 
11  return  $\pi$ 
```

Knuth-Morris-Pratt(KMP) algorithm

Time complexity

Running time of compute-prefix-function is $\theta(m)$.
The matching time of KMP-Matcher is $\theta(n)$.

Question: Consider text and pattern as following:-

T = bacbababaabcbab

P = aba

Find all valid shifts using KMP algo.

Question: Compute the prefix function for the pattern ababbabbabbabbabb.

AKTU Examination Questions

1. Write an algorithm for Naïve string matcher.
2. Write KMP algorithm for string matching. Perform the KMP algorithm to search the occurrences of the pattern abaab in the text string abbabaabaabab.
3. Write Rabin Karp string matching algorithm. Working modulo $q=11$, how many spurious hits does the Rabin karp matcher in the text $T= 3141592653589793$, when looking for the pattern $P=26$.
4. Explain and Write the Knuth-Morris-Pratt algorithm for pattern matching also write its time complexity.
5. Describe in detail Knuth-Morris-Pratt string matching algorithm. Compute the prefix function π for the pattern ababbabbabbababbabb when the alphabet is $\Sigma = \{a,b\}$.
6. Compute the prefix function π for the pattern $P= a b a c a b$ using KNUTHMORRIS –PR