SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A+' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF INFORMATION TECHNOLOGY

19ITB201– Design and Analysis of Algorithms

II YEAR IV SEM

UNIT 4 – FLOW NETWORKS AND STRING MATCHING

TOPIC – String Matching-Navie String Matching Algorithm

NAIVE STRING MATCHING

What is a String?

In C, String is a sequence of characters or more specifically can be regarded as an array of characters.

String Matching

One of the most commonest operation done with strings is String Matching.

WHAT IS STRING MATCHING?

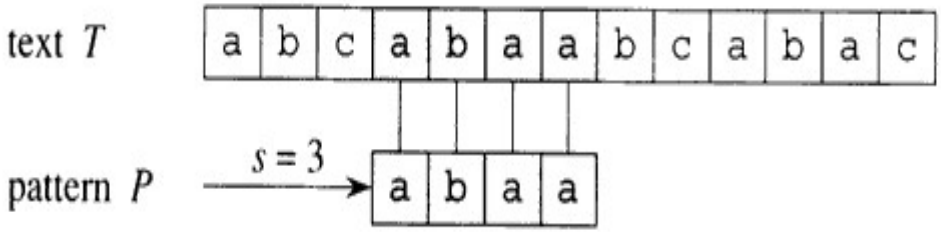
If we are given a string “p” which is nothing but stream of characters p[0],p[1],p[2]……..p[n-1] (where n is the length of the string),then searching for a specific pattern in the string is known as string matching

Applications of String Matching

- Intrusion Detection
- String matching in bioinformatics
- String matching in Digital Forensics

An example of string mat ching

Here we discuss an example on string matching where the given text consists of a pattern which is to be searched for. The following figure shows the location of pattern P in a given text T.



Algorithm For String Matching

There are many ways to do the String Matching but here we discuss the concept of

Naïve String Matching.

Naive string matching

The naive algorithm finds all valid shifts using a loop that checks the condition $P[1..m] = T[s+1..s+m]$ for each of the $n-m+1$ possible values of s.

```
NAIVE-STRING-MATCHER(T, P)
1  n ← length[T]
2  m ← length[P]
3  for s ← 0 to n − m
4        do if P[1..m] = T[s + 1..s + m]
5              then print "Pattern occurs with shift" s
```

Running time: O((n-m+1)m).

Problem with naive algorithm

- Suppose p=ababc, T=cababababcd

- T: c a b a b a b c d
- P: a ...
- P: a b a b c
- P: a ...
- P: a b a b c
- Whenever a character mismatch occurs after matching of several characters, the comparison begins by going back in T from the character which follows the last beginning character. This makes the process very slow.

Solution to the Problem

•There are many more algorithms which work in a more efficient way than the naive string matching and are listed below:-

- Rabin–Karp string search algorithm Finite-state automaton Knuth–Morris–Pratt algorithm Boyer–Moore string search algorithm

THANK YOU