

Report

Introduction:

To analyse and calculate the timing of substring search for four methods:

- 1) Brute force algorithm
- 2) Knuth - Morris - Pratt algorithm
- 3) Boyer - Moore algorithm
- 4) Rabin Karp algorithm

Given substring is “it is far far better thing that i do than i have ever done” in the text of Two cities (tale.txt).

Intuitive thinking:

I think Rabin Karp algorithm is the fastest algorithm to find the substring search because it compares by matching the hash value of the pattern with the hash value of current substring of text and its worst case time complexity is $O(NM)$.

Also, I think that Brute force method is slower because we have to check every single character from the text to match against the pattern. The time complexity is $O(N*M)$.

Procedure:

- 1) We consider the “tale.txt” file to search the “it is far far better thing that i do than i have ever done” substring.
- 2) For this, we implement code to find the running time for each four algorithms i.e., Brute force, KMP, Boyer-Moore and Rabin-Karp.
- 3) I observed the time taken to search the pattern in four algorithms and found the required time in milliseconds.

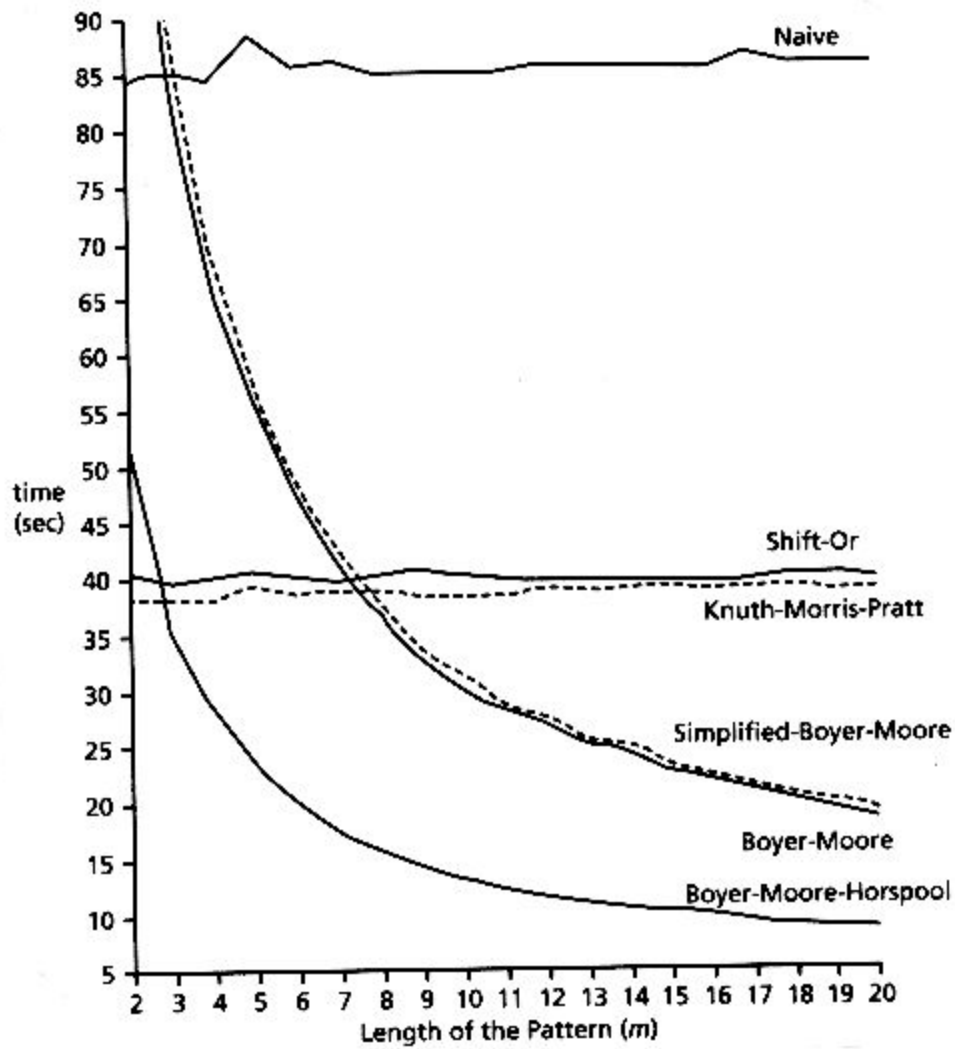
Observation table:

| S.no | Algorithm | Time taken(Milliseconds) |
|------|-------------|--------------------------|
| 1. | Brute force | 55 |
| 2. | KMP | 52 |
| 3. | Boyer Moore | 38 |
| 4. | Rabin Karp | 273 |

```
C:\Windows\System32\cmd.exe
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>java RabinKarp
Pattern found!!!
It is a far, far better thing that I do, than I have ever done
It took 273 milliseconds
273
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>javac BoyerMoore.java
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>java BoyerMoore
Pattern Found!!!
It is a far, far better thing that I do, than I have ever done
It took 38 milliseconds
38
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>javac Brute.java
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>java Brute
Pattern Found
It is a far, far better thing that I do, than I have ever done
It took 55 milliseconds
55
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>javac KMP.java
C:\Users\prasamsa\Desktop\ADS-2\ADS-2-assignments\m23>java KMP
Pattern Found!!!
It is a far, far better thing that I do, than I have ever done
It took 52 milliseconds
52
```

Analysis:

- As we observe the table readings from the observation table, we come to conclusion that Boyer Moore algorithm takes lesser time than any other algorithm for substring search.
- The reason for that could be “it slides the pattern over the text one by one. KMP algorithm does preprocessing over the pattern so that the pattern can be shifted by more than one. Unlike the previous pattern searching algorithms, Boyer Moore algorithm starts matching from the last character of the pattern.
- We also came to know that Brute force algorithm is the slowest algorithm for substring search as we already assumed.
- In KMP algorithm, more clearly we focus on sub-strings of patterns that are either prefix and suffix. It uses DFA simulation and takes average time to find the substring.



Conclusion:

Hence, I would like to conclude that Boyer Moore is the fastest algorithm and Brute force algorithm is the slowest one for substring search.