

## A Comparison between string-matching algorithms – Task 2

Mohamed Samine

May 15, 2021

1. Sunday is at least twice as fast as Brute-force.

Sunday algorithm can perform at least twice as fast as brute force in the case where the text does not include so many characters that are in the pattern which allow Sunday to do many shifts that equal the length of the pattern in case of mismatches, Sunday also perform better than brute force for large text sizes with few occurrences of the pattern.

### Pattern example: *EasterEgg*

Text example:

[illegible]

Result Example:

"pattern", "bruteForce", "sunday", "rabinKarp", "fsm", "kmp"

EasterEgg, 0.03965879994630814, 0.015965600296854975, 0, 0, 0

2. Sunday is at least twice as fast as KMP.

As already pointed out before Sunday algorithms has the advantage of doing large shifts that equals the length of the pattern when the next characters outside the current comparison window is not included in the pattern, as for KMP it has the property of avoiding to re-compare the already matched characters in the pattern.

To achieve our goal of that is Sunday to be at least twice as fast as KMP we take a pattern where there are no repeated characters so that KMP does not find any prefix/suffix in the pattern, and we also make sure that the text is composed of not many characters that are included in the patterns, by doing so we give Sunday algorithm more chances to do the pattern length shifts more often.

Pattern example: Eastrg

Text example:

[illegible]

Result Example:

"pattern", "bruteForce", "sunday", "rabinKarp", "fsm", "kmp"

Eastrg, 0, 0.02501129971444607, 0, 0, 0.08048310001194477

### 3. KMP is at least twice as fast as Rabin-Karp

We know that KMP algorithm does not go back and compare the pattern from the beginning, we also know that Rabin Karp performs bad with texts that has many repetitive characters, as it must double check every potential match which boils down to doing brute force plus the overhead of computing the hash for every new character introduced from the text.

So, to achieve the situation given, we can simply take a text of repetitive characters and a pattern of repetitive characters.

Pattern example: aaaa

Text example:

```
aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
```

Result example:

"pattern","bruteForce","sunday","rabinKarp","fsm","kmp"

aaaa, 0, 0, 0.023652300491929054, 0, 0.00966650006175041

### 4. Rabin-Karp is at least twice as fast as Sunday.

As we know that for **Sunday** algorithm to be able to do large shifts more often the text should not have many characters that appear in the pattern, we also know that **Sunday** has preprocessing phase that is directly proportional to *pattern length*, during this phase it prepares the precomputed ascii table based on the given pattern.

Additionally, the **Rabin-Karp** algorithm has a space complexity of  $O(1)$  and has no preprocessing phase.

With the three points in mind, the situation can be achieved by setting the text to be consisting of the same repetitive character, and the pattern to be long and made of the same repetitive character.

This will result on **Sunday** taking more time in the preprocessing stage and making shifts of one character at time which will give the **Rabin-Karp** algorithm advantage over the **Sunday** algorithm.

Pattern example:

[illegible]

Text example:

[illegible]

Result example:

```
"bruteForce","sunday","rabinKarp","fsm","kmp"
```

0, 1.34822939966619, 0.14870780016481877, 0, 0