Boyer-Moore Strategy

We visited the links provided by Harun Yahya Ozturk

On his github repository, And

https://www.geeksforgeeks.org/dsa/boyer-moore-algorithm-for-pattern-searching/ (https://www.geeksforgeeks.org/dsa/boyer-moore-algorithm-for-pattern-searching/)

https://www.topcoder.com/thrive/articles/boyer-moore-algorithm-with-bad-character-heuristic (https://www.topcoder.com/thrive/articles/boyer-moore-algorithm-with-bad-character-heuristic)

Then we asked ChatGPT to explain them to us.

Under ChatGPT's guidance we were able to implement

A functional Boyer-Moore Algorithm using both bad character and good suffix rules.

PreAnalysis Strategy

Pattern length ≤ 3 → Naive

Reason: These patterns lead to very small m.

For small m, Naive often outperforms everything

Patterns with repeating prefixes → KMP

Reason: KMP is built for periodic or repetitive patterns,

Because its LPS table eliminates redundant comparisons.

Long pattern (m > 10) and long text (n > 1000) → Rabin–Karp

Reason: Hashing becomes beneficial when both text and pattern are long.

Everything else → Boyer–Moore

Reason: This is the general-purpose best-case algorithm.

It performs very well when,

pattern is moderately large,

text is mixed/randomized,

and the alphabet is not extremely small.

There are some other checks for certain edge cases.

They are explained at the comments within the method.

Result Analysis

Naive wins almost every case where:

Pattern is short (m < ~10)

Text is relatively small

Few matches are expected

No special structure

KMP wins when:

Pattern is periodic, repeated (examples: "aaaaa...", "abababab")

Single-character pattern cases

DNA sequence case (very small alphabet!)

Alternating patterns

Boyer–Moore wins when:

Text is very large

Pattern is longer

Alphabet is large or random-ish

Rabin–Karp wins when:

Hashing quickly filters mismatches

Cases like:

Simple match

Worst case for Naive

Best case for BM

KMP advantage case (surprisingly)

Palindrome patterns

Empty text

RK seems to benefit from its low overhead + fast fail checks.

Our Student Information

Hasan Tarık BAKAR 22050111007

Mehmet Emre YILDIZ 22050111065