



Busca em cadeias (Força bruta e Rabin-Karp)

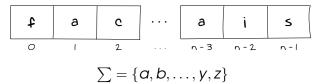
Projeto e Análise de Algoritmos

Bruno Prado

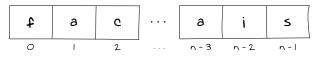
Departamento de Computação / UFS

- ▶ O que é uma cadeia?
 - \blacktriangleright É uma sequência de símbolos T com tamanho n

- O que é uma cadeia?
 - \blacktriangleright É uma sequência de símbolos T com tamanho n
 - lacktriangle Os símbolos são definidos por um alfabeto finito \sum



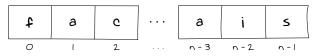
- O que é uma cadeia?
 - \blacktriangleright É uma sequência de símbolos T com tamanho n
 - lacktriangle Os símbolos são definidos por um alfabeto finito \sum



$$\sum = \{a, b, \dots, y, z\}$$

- Aplicações multidisciplinares
 - Biologia: representação da cadeia de DNA, sendo composta pelos símbolos A, C, G, T

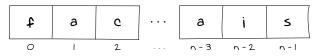
- O que é uma cadeia?
 - \blacktriangleright É uma sequência de símbolos T com tamanho n
 - lacktriangle Os símbolos são definidos por um alfabeto finito \sum



$$\sum = \{a, b, \dots, y, z\}$$

- Aplicações multidisciplinares
 - Biologia: representação da cadeia de DNA, sendo composta pelos símbolos A, C, G, T
 - Computação: armazenamento de texto através do tipo string, com o padrão de codificação ASCII

- O que é uma cadeia?
 - \blacktriangleright É uma sequência de símbolos T com tamanho n
 - lacktriangle Os símbolos são definidos por um alfabeto finito \sum



$$\sum = \{a, b, \dots, y, z\}$$

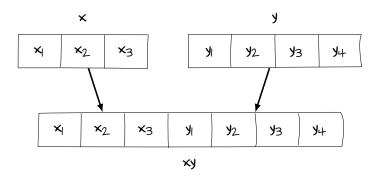
- Aplicações multidisciplinares
 - Biologia: representação da cadeia de DNA, sendo composta pelos símbolos A, C, G, T
 - Computação: armazenamento de texto através do tipo string, com o padrão de codificação ASCII

- Notação e terminologia
 - Todos as cadeias de tamanho finito que podem ser construídas do alfabeto finito ∑ é definido por ∑*

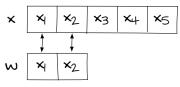
- Notação e terminologia
 - Todos as cadeias de tamanho finito que podem ser construídas do alfabeto finito ∑ é definido por ∑*
 - ightharpoonup Uma cadeia vazia é denotada pelo símbolo arepsilon

- Notação e terminologia
 - Todos as cadeias de tamanho finito que podem ser construídas do alfabeto finito ∑ é definido por ∑*
 - lacktriangle Uma cadeia vazia é denotada pelo símbolo arepsilon
 - O tamanho de uma cadeia x é definida por |x|

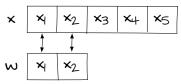
- Notação e terminologia
 - Todos as cadeias de tamanho finito que podem ser construídas do alfabeto finito ∑ é definido por ∑*
 - lacktriangle Uma cadeia vazia é denotada pelo símbolo arepsilon
 - ightharpoonup O tamanho de uma cadeia x é definida por |x|
 - A concatenação de duas cadeias x e y resulta em uma cadeia xy com os caracteres de x seguidos dos caracteres de y, com tamanho total de |x| + |y|



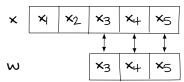
- Notação e terminologia
 - Prefixo: a cadeia w é um prefixo da cadeia x ($w \sqsubset x$) se x = wy, com $y \in \sum^* e |w| \le |x|$



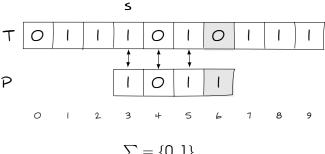
- Notação e terminologia
 - Prefixo: a cadeia w é um prefixo da cadeia x ($w \sqsubset x$) se x = wy, com $y \in \sum^* e |w| \le |x|$



Sufixo: a cadeia w é sufixo da cadeia x ($w \supset x$) se x = yw, com $y \in \sum^* e |w| \le |x|$



- Como pode ser definida a busca em cadeias?
 - É o processo para encontrar todas as ocorrências de um padrão P em uma cadeia T que possuem m e n símbolos, respectivamente, onde m ≤ n

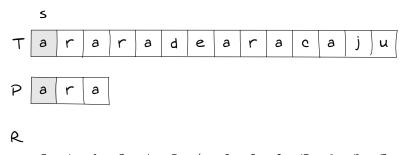


$$\sum_{\substack{P | P = m = 4, |T| = n = 10 \\ 0 \le s \le n - m}} \{0, 1\}$$

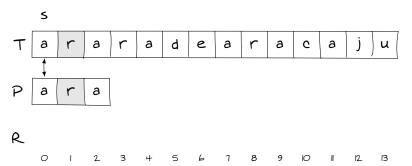
Força bruta

```
// Padrão de tipos por tamanho
  #include <stdint.h>
   // Busca em cadeias por força bruta
   void forca_bruta(int32_t* R, char* T, char* P) {
       // Tamanhos das cadeias T e P
6
       int32_t n = strlen(T), m = strlen(P);
       // Iterando na cadeia T até o índice n - m
       for(int32_t s = 0; s <= n - m; s++) {
           // Contador de diferenças
9
           uint32_t d = 0;
10
           // Iterando na cadeia P enquanto d for zero
11
           for(int32_t i = 0; i < m && d == 0; i++) {
12
               // Caso os símbolos sejam diferentes
13
               if(P[i] != T[s + i]) d++;
14
15
           // O índice da combinação é armazenado
16
           if (d == 0) inserir(R, s);
17
18
19
```

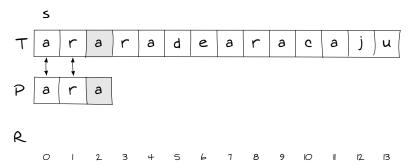
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



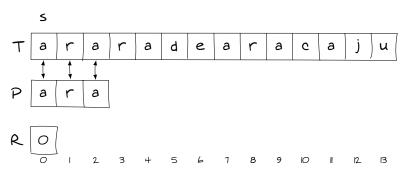
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



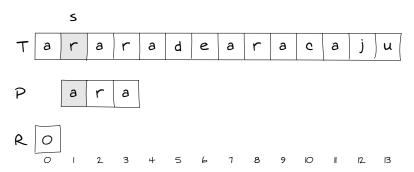
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



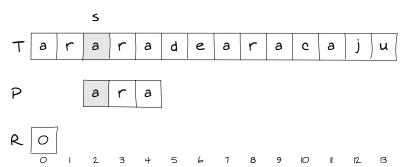
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



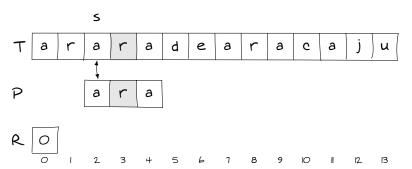
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



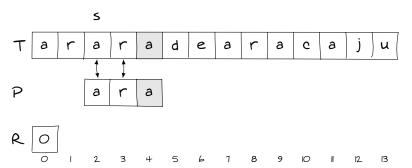
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



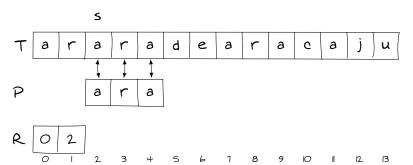
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



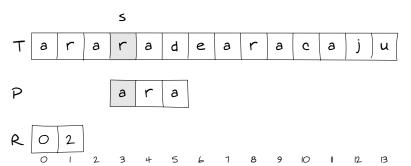
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



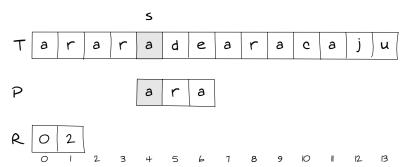
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



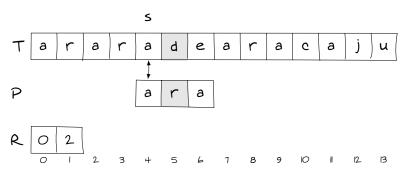
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



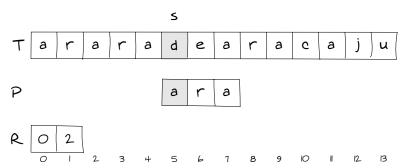
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



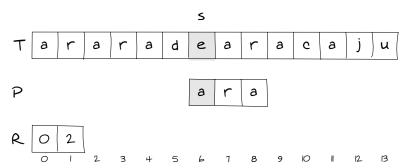
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



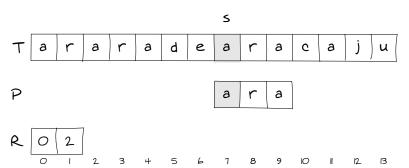
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



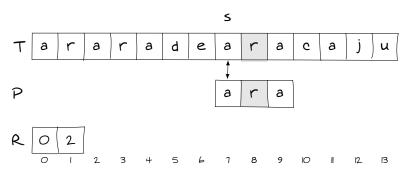
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



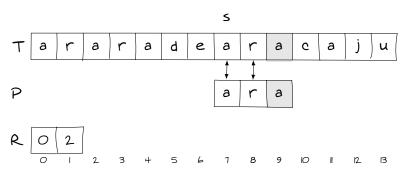
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



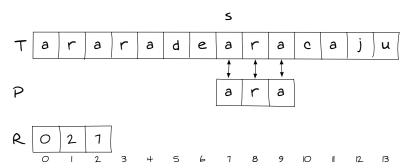
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



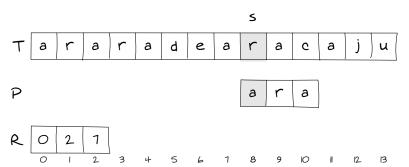
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



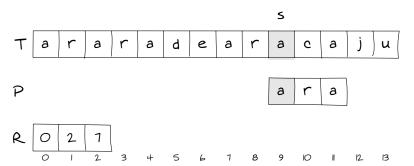
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



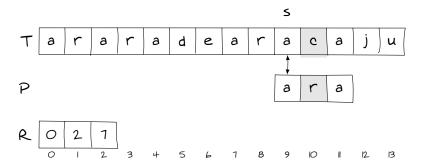
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



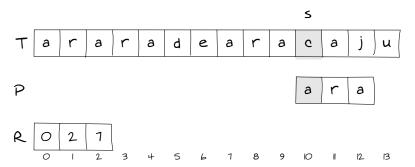
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



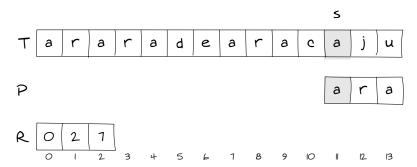
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



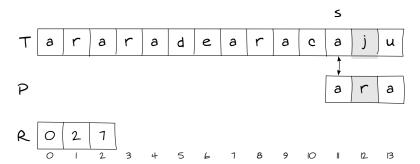
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



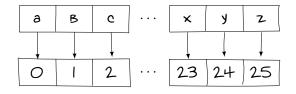
- Força bruta
 - Realizando a busca pelo padrão P = ara na cadeia T = araradearacaju através do seu deslocamento e comparação um para um dos símbolos das cadeias



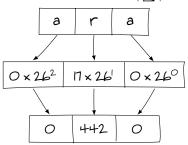
- ► Força bruta
 - Análise de complexidade
 - Espaço O(n+m)
 - ► Tempo $\Omega(n-m+1)$ e $O((n-m+1)\times m)$

- Rabin-Karp
 - A busca na cadeia de símbolos é feita com o pré-processamento do padrão de texto

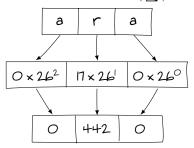
- Rabin-Karp
 - A busca na cadeia de símbolos é feita com o pré-processamento do padrão de texto
 - Considerando um alfabeto $\sum = \{a, b, ..., y, z\}$ com 26 símbolos, as cadeias podem ser representadas como números em base $b = |\sum| = 26$



- Rabin-Karp
 - A busca na cadeia de símbolos é feita com o pré-processamento do padrão de texto
 - Considerando um alfabeto $\Sigma = \{a, b, ..., y, z\}$ com 26 símbolos, as cadeias podem ser representadas como números em base $b = |\Sigma| = 26$



- Rabin-Karp
 - A busca na cadeia de símbolos é feita com o pré-processamento do padrão de texto
 - Considerando um alfabeto $\Sigma = \{a, b, ..., y, z\}$ com 26 símbolos, as cadeias podem ser representadas como números em base $b = |\Sigma| = 26$



$$P = ara \rightarrow p = 0 + 442 + 0 = 442$$

- Rabin-Karp
 - A comparação das cadeias é feita comparando as representações numéricas do texto t e do padrão p

- Rabin-Karp
 - A comparação das cadeias é feita comparando as representações numéricas do texto t e do padrão p
 - Como o alfabeto e o padrão de busca podem conter muitos símbolos, o valor numérico b^m − 1 pode não ser representável na plataforma (2⁶⁴ − 1 ≤ b^m − 1)

- Rabin-Karp
 - A comparação das cadeias é feita comparando as representações numéricas do texto t e do padrão p
 - Como o alfabeto e o padrão de busca podem conter muitos símbolos, o valor numérico b^m − 1 pode não ser representável na plataforma (2⁶⁴ − 1 ≤ b^m − 1)
 - Para contornar esta limitação, é aplicada a operação de módulo com um número primo q para restringir o tamanho do número gerado p mod q e reduzir a possibilidade de colisões

- Rabin-Karp
 - A comparação das cadeias é feita comparando as representações numéricas do texto t e do padrão p
 - Como o alfabeto e o padrão de busca podem conter muitos símbolos, o valor numérico b^m − 1 pode não ser representável na plataforma (2⁶⁴ − 1 ≤ b^m − 1)
 - Para contornar esta limitação, é aplicada a operação de módulo com um número primo q para restringir o tamanho do número gerado p mod q e reduzir a possibilidade de colisões

O valor de $b \times q$ precisa ser representável ($b \times q \le 2^{64} - 1$)

Rabin-Karp

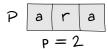
```
// Padrão de tipos por tamanho
   #include <stdint.h>
   // Busca em cadeias com Rabin-Karp
   void rabin_karp(int32_t* R, char* T, char* P,
      uint32_t b, uint32_t q) {
       // Variáveis auxiliares
5
       int32_t n = strlen(T), m = strlen(P), h = pow(b,
          m - 1) \% q, t = 0, p = 0;
       // Pré-processamento numérico das cadeias p e t
       for(int32_t i = 0; i < m; i++) {
           p = (b * p + v(P[i])) % q;
           t = (b * t + v(T[i])) % q;
10
11
21
```

Rabin-Karp

```
// Padrão de tipos por tamanho
  #include <stdint.h>
   // Busca em cadeias com Rabin-Karp
   void rabin_karp(int32_t* R, char* T, char* P,
      uint32_t b, uint32_t q) {
       // Variáveis auxiliares
5
       int32_t n = strlen(T), m = strlen(P), h = pow(b,
           m - 1) \% q, t = 0, p = 0;
       // Iterando na cadeia T até o índice n - m
12
       for(int32_t s = 0; s \le n - m; s++) {
1.3
           // Comparando o padrão com a cadeia
14
           if((p == t) && igual(P, m, T, s))
15
               // Armazenando índice de combinação
16
                inserir(R, s);
17
           // Deslocamento numérico
18
           t = (b * (t - v(T[s]) * h) + v(T[s + m])) % q;
19
20
21
```

- Rabin-Karp
 - Cada subcadeia de T de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

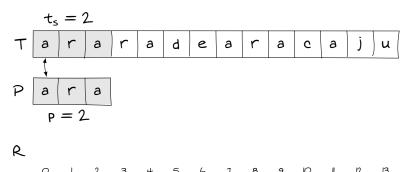
$$t_s = 2$$
Tararadearacaju



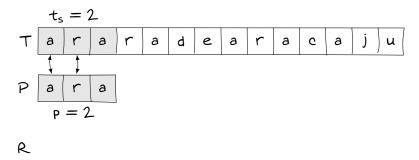
R

0 1 2 3 4 5 6 1 8 9 10 11 12 13

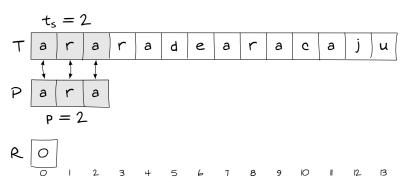
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

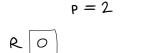


- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

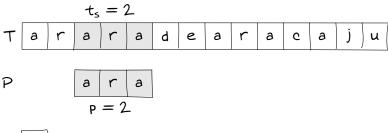


- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

$$t_s = 4$$
Tararadearacaju



- Rabin-Karp
 - ► Cada subcadeia de T de tamanho m tem seu valor numérico t_s calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2





3

4 5

5

6

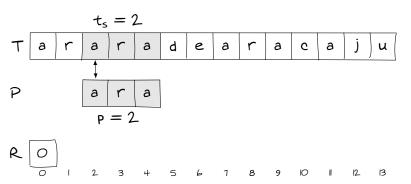
1

3

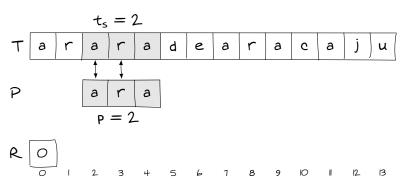
ı

12

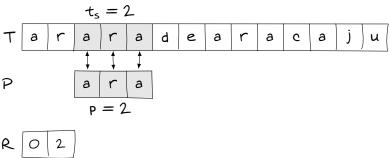
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



- Rabin-Karp
 - Cada subcadeia de T de tamanho m tem seu valor numérico t_s calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



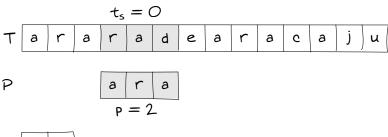
- Rabin-Karp
 - Cada subcadeia de T de tamanho m tem seu valor numérico t_s calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2







- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



3 4

+ 5

6

ı

12

- Rabin-Karp
 - Cada subcadeia de T de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

3

4 5

Ь

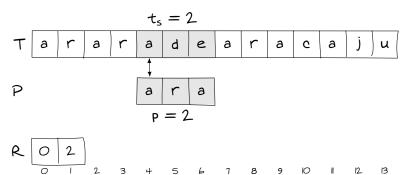
.

8

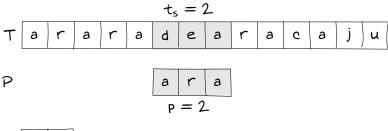
ll

12

- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

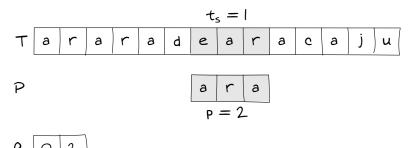


- Rabin-Karp
 - Cada subcadeia de T de tamanho m tem seu valor numérico t_s calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2

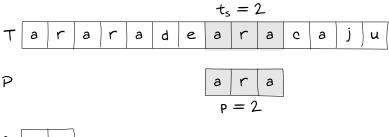




- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



3

4

5

6

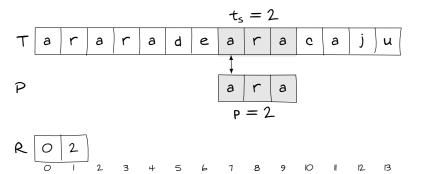
1

3

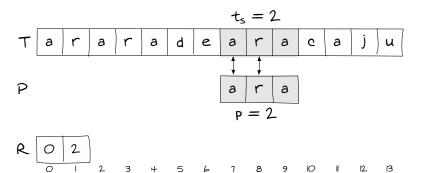
12

2_

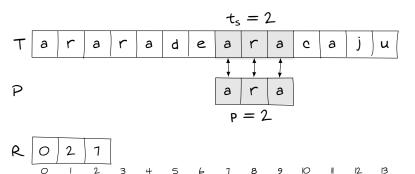
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



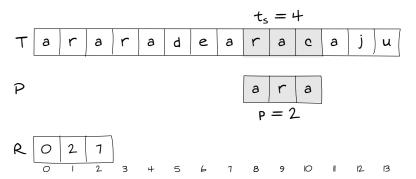
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



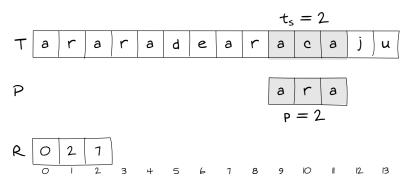
- Rabin-Karp
 - Cada subcadeia de T de tamanho m tem seu valor numérico t_s calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



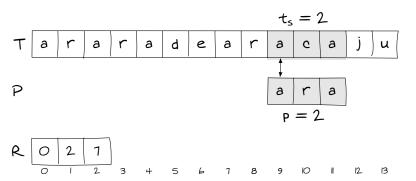
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



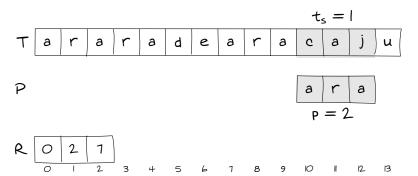
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



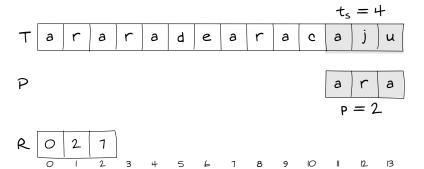
- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



- Rabin-Karp
 - Cada subcadeia de 7 de tamanho m tem seu valor numérico t₅ calculado e comparado com o valor de p, considerando b = 26, q = 5 e p = 442 mod 5 = 2



- Rabin-Karp
 - ▶ Análise de complexidade
 - Espaço O(n+m)
 - ► Tempo $\Theta(m) + \Omega(n-m+1)$ e $\Theta(m) + O((n-m+1) \times m)$

Exemplo

- - Execute passo a passo a busca na cadeia
 - Faça uma análise comparativa dos algoritmos, destacando seus princípios de funcionamento, além de suas respectivas vantagens e desvantagens