



Projeto e Análise de Algoritmos

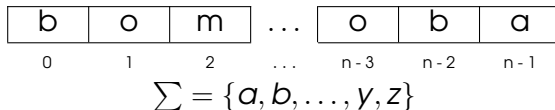
Busca em cadeias (Força bruta e RK)

Bruno Prado

Departamento de Computação / UFS

Introdução

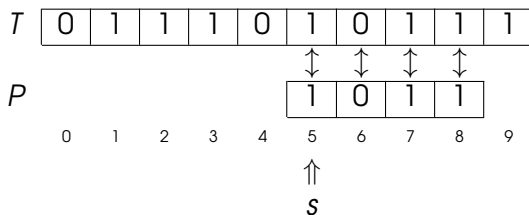
- ▶ O que é uma cadeia?
 - ▶ É uma sequência de símbolos T com tamanho n
 - ▶ Os símbolos são definidos por um alfabeto finito Σ



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

Introdução

- ▶ O que é uma busca em cadeia?
 - ▶ É o processo para encontrar todas as ocorrências de um padrão em uma cadeia T que possui n símbolos
 - ▶ Para a busca é utilizada a cadeia de padrão P com quantidade de símbolos $m \leq n$
 - ▶ As cadeias P e T utilizam um alfabeto finito Σ

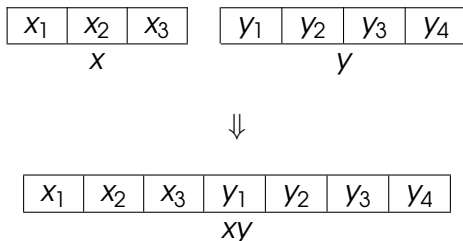


$$\begin{aligned}\Sigma &= \{0, 1\} \\ |T| &= n = 10, |P| = m = 4 \\ 0 &\leq s \leq n - m\end{aligned}$$

Introdução

► Notação e terminologia

- É definido por Σ^* todos os conjuntos de cadeias de tamanho finito que podem ser construídas do alfabeto finito Σ
- Uma cadeia vazia é denotada pelo símbolo ε
- 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 $|x| + |y|$

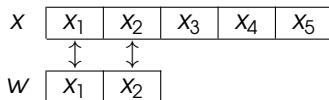


Introdução

► Notação e terminologia

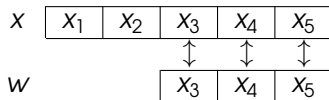
► Prefixo

- A cadeia w é um prefixo da cadeia x se $x = wy$, para alguma cadeia $y \in \Sigma^*$
- Denotado por $w \sqsubset x$, com $|w| \leq |x|$



► Sufixo

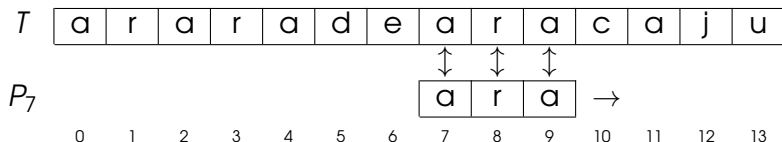
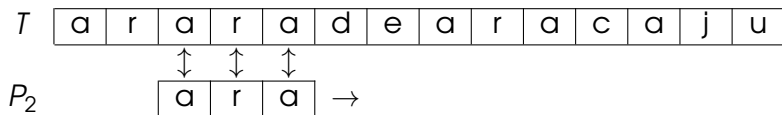
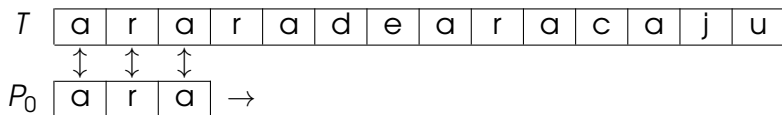
- A cadeia w é sufixo da cadeia x se $x = yw$, para alguma cadeia $y \in \Sigma^*$
- Denotado por $w \sqsupset x$, com $|w| \leq |x|$



Busca em cadeias

► Força bruta

- Realiza a busca por um padrão na cadeia através do seu deslocamento e comparação dos símbolos
- Texto $T = \text{"araradearacaju"}$ e padrão $P = \text{"ara"}$



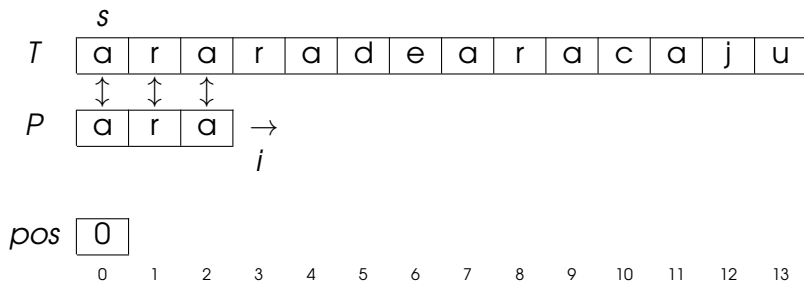
Busca em cadeias

- Força bruta
 - Implementação em C

```
void busca_fb(int pos(), char T(), char P()) {  
    int i, j, s;  
    unsigned int n = strlen(T);  
    unsigned int m = strlen(P);  
    for(s = 0; s <= n - m; s++) {  
        for(i = 0, j = 0; i < m && j == 0; i++) {  
            if(P(i) != T(s + i))  
                j++;  
        }  
        if(j == 0)  
            inserir(pos, s);  
    }  
}
```

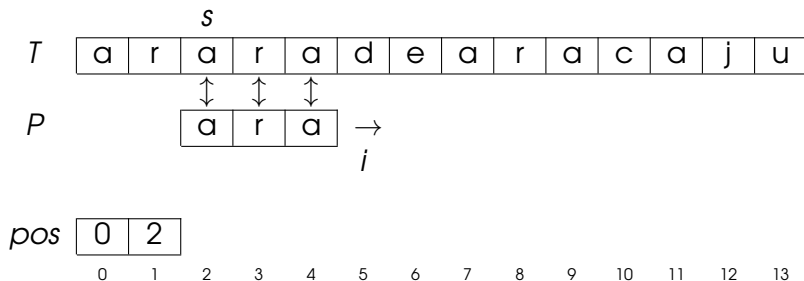
Busca em cadeias

- ▶ Força bruta
 - ▶ Execução do algoritmo



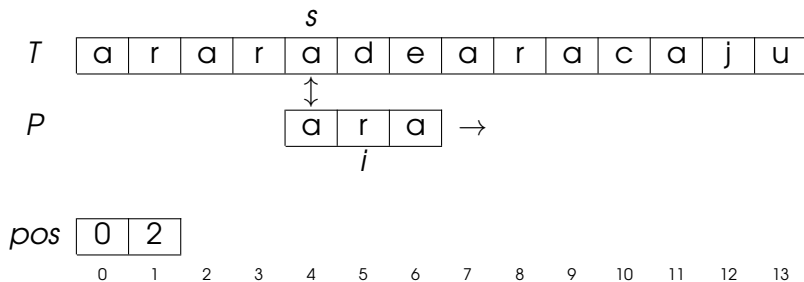
Busca em cadeias

- ▶ Força bruta
 - ▶ Execução do algoritmo



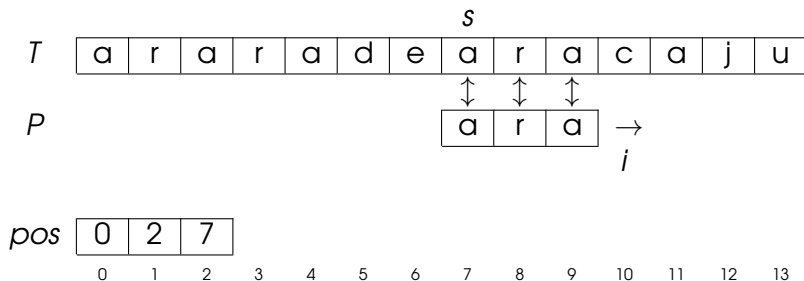
Busca em cadeias

- ▶ Força bruta
 - ▶ Execução do algoritmo



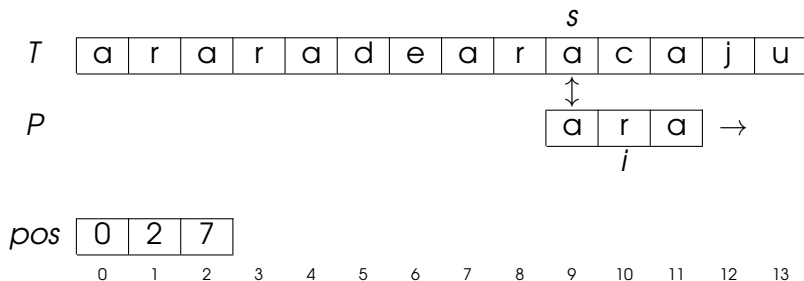
Busca em cadeias

- ▶ Força bruta
 - ▶ Execução do algoritmo



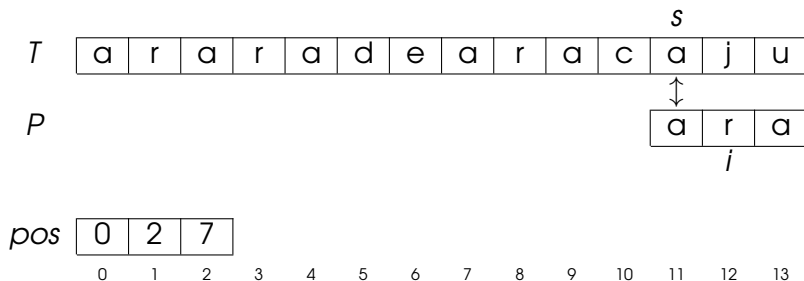
Busca em cadeias

- ▶ Força bruta
 - ▶ Execução do algoritmo



Busca em cadeias

- ▶ Força bruta
 - ▶ Execução do algoritmo

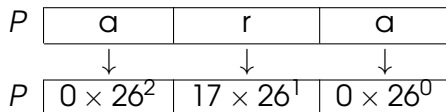
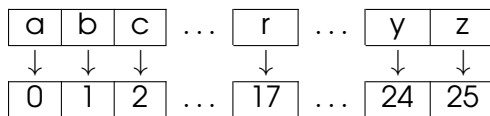


Busca em cadeias

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

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ A busca na cadeia de símbolos é feita com o pré-processamento do padrão de texto
 - ▶ O alfabeto $\Sigma = \{a, b, \dots, y, z\}$ com 26 símbolos e as cadeias podem ser vistas como números representados em base $b = |\Sigma| = 26$



$$\downarrow$$
$$p = 0 + 442 + 0 = 442$$

Cadeia $P \rightarrow$ Número p

Busca em cadeias

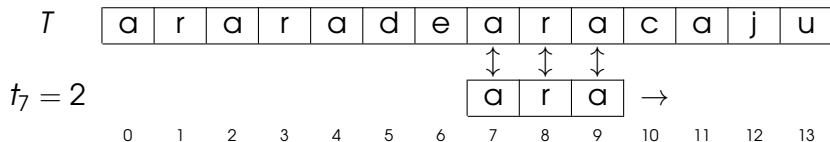
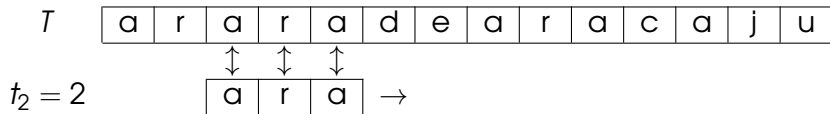
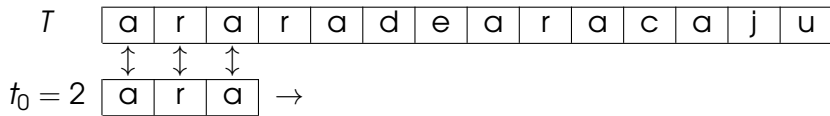
- ▶ Rabin-Karp (RK)
 - ▶ A comparação das cadeias é feita comparando as representações numéricas do texto t e do padrão p
 - ▶ Como o alfabeto e o padrão de busca pode conter muitos símbolos, o valor numérico pode ser muito grande para ser representado
 - ▶ É aplicada a operação de módulo utilizando um número primo q para armazenar o número gerado, atendendo a restrição de que o valor de $b \times q$ é representável em uma palavra do sistema

$$b \times q \leq 2^{32} - 1$$

Busca em cadeias

► Rabin-Karp (RK)

- Cada subcadeia de T de tamanho m tem seu valor numérico t_i calculado e comparado com o valor de p
- Com $b = 26$ e $q = 5$, o valor de $p = 442 \bmod 5 = 2$



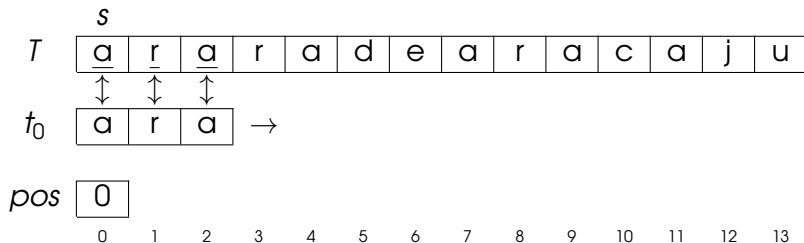
Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Implementação em C

```
void busca_rk(int pos(), char T(), char P(), int b, int q) {  
    unsigned int n = strlen(T);  
    unsigned int m = strlen(P);  
    unsigned int i, s, h = pow(b, m - 1) % q, t = 0, p = 0;  
    for(i = 0; i < m; i++) {  
        p = (b * p + v(P(i))) % q;  
        t = (b * t + v(T(i))) % q;  
    }  
    for(s = 0; s <= n - m; s++) {  
        if((p == t) && igual(P, m, T, s))  
            inserir(pos, s);  
        t = (b * (t - v(T(s)) * h) + v(T(s + m))) % q;  
    }  
}
```

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo

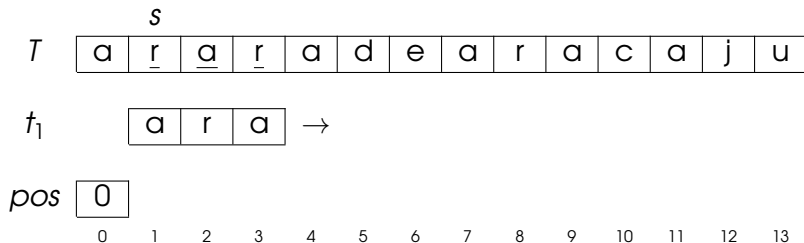


$$p = 2$$

$$\begin{aligned} t &= (0 \times 26^2 + 17 \times 26^1 + 0 \times 26^0) \bmod 5 \\ &= 442 \bmod 5 \\ &= 2 \end{aligned}$$

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo



$$p = 2$$

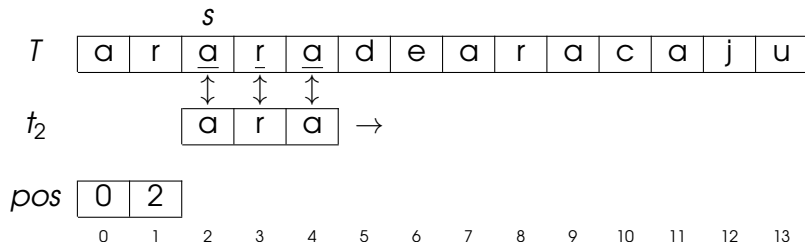
$$t = 17 \times 26^2 + 0 \times 26^1 + 17 \times 26^0 \bmod 5$$

$$= 11509 \bmod 5$$

$$= 4$$

Busca em cadeias

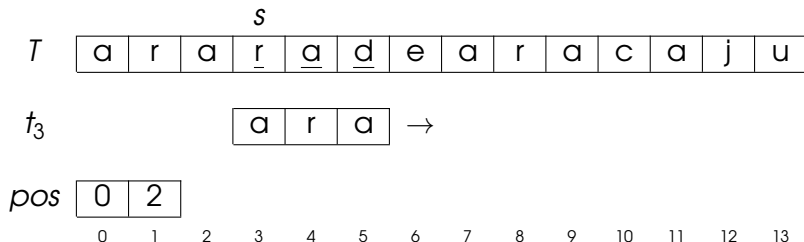
- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo



$$\begin{aligned}p &= 2 \\t &= (0 \times 26^2 + 17 \times 26^1 + 0 \times 26^0) \bmod 5 \\&= 442 \bmod 5 \\&= 2\end{aligned}$$

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo

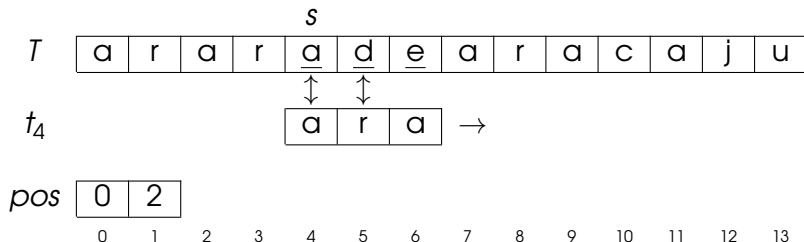


$$p = 2$$

$$\begin{aligned} t &= (17 \times 26^2 + 0 \times 26^1 + 3 \times 26^0) \bmod 5 \\ &= 11495 \bmod 5 \\ &= 0 \end{aligned}$$

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo

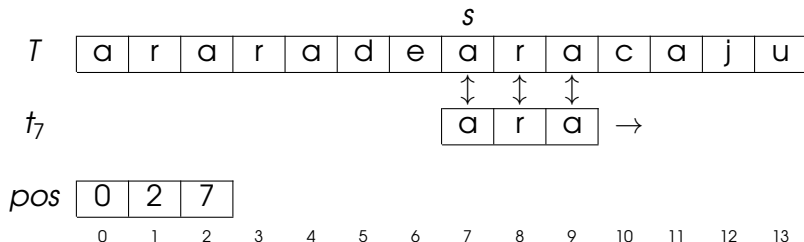


$$p = 2$$

$$\begin{aligned} t &= (0 \times 26^2 + 3 \times 26^1 + 4 \times 26^0) \bmod 5 \\ &= 82 \bmod 5 \\ &= 2 \end{aligned}$$

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo

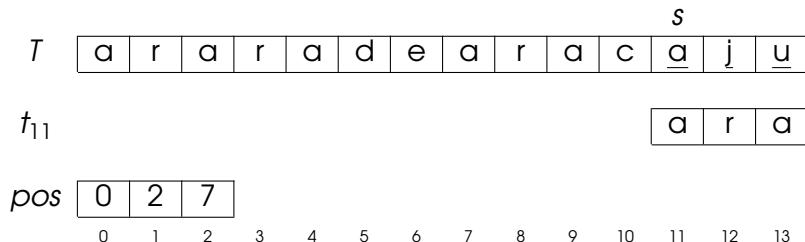


$$p = 2$$

$$\begin{aligned} t &= (0 \times 26^2 + 17 \times 26^1 + 0 \times 26^0) \bmod 5 \\ &= 442 \bmod 5 \\ &= 2 \end{aligned}$$

Busca em cadeias

- ▶ Rabin-Karp (RK)
 - ▶ Execução do algoritmo



$$\begin{aligned}p &= 2 \\t &= (0 \times 26^2 + 9 \times 26^1 + 20 \times 26^0) \bmod 5 \\&= 254 \bmod 5 \\&= 4\end{aligned}$$

Busca em cadeias

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

Exemplo

- ▶ Aplique os algoritmos de busca em cadeias para encontrar o padrão "111000" na sequência binária "10111000110111100010101100011100001101101111"
- ▶ Execute passo a passo a busca na cadeia
- ▶ Faça uma análise comparativa dos algoritmos
 - ▶ Princípio de funcionamento
 - ▶ Vantagens e desvantagens