

Algoritmos e Estruturas de Dados III

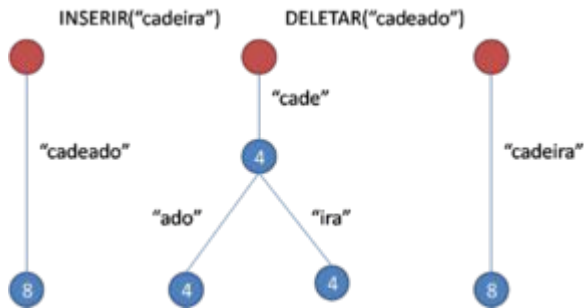
10.1 Casamento de Padrões

Prof. Hayala Curto
2022



PUC Minas

Roteiro do Conteúdo



Revisão de Pesquisa Digital

- Trie
- Patrícia

Casamento de Padrões

- Introdução
- Força Bruta

○

Revisão Pesquisa Digital



PUC Minas

Casamento de Padrões

Motivação

Suponha que se queira armazenar um texto e, em seguida, tentar localizar frases neste texto.

Características

- Chaves não são tratadas como um único elemento indivisível
- Não há suposição alguma sobre o tamanho da chave
- Chaves de tamanhos variáveis
- Chaves são conj. de caracteres ou dígitos de um alfabeto
- Chaves são comparadas dígito a dígito (ou caracter a caracter)

Pesquisa Digital - Trie

Árvore Digital (Trie - Fredkin/60), do inglês "re**Tri**eval"

Seja $\mathbf{S} = \{ s_1, \dots, s_n \}$ um conj. de n chaves formadas por uma sequência de elem. \mathbf{d}_j denominados dígitos

Seja Σ o alfabeto com m dígitos

Seja \mathbf{T} uma árvore digital então :

- \mathbf{T} é uma árvore m -ária
- Se um nó \mathbf{v} é o j -ésimo filho de seu pai, então \mathbf{v} corresponde ao dígito \mathbf{d}_j de Σ $1 \leq j \leq m$
- Para cada nó \mathbf{v} , a sequência de dígitos definida pelo caminho desde a raiz de \mathbf{T} até \mathbf{v} corresponde a um prefixo de alguma chave de \mathbf{S} .
- A raiz de \mathbf{T} sempre existe e não corresponde a nenhum dígito.

Pesquisa Digital

- Baseada na representação das chaves como uma sequência de caracteres ou dígitos.
- Parecido com busca em dicionários com "índices de dedo"
- Vantajosa quando chaves são grandes e tem tamanho variável
- Localização de todas as ocorrências de determinada cadeia em um texto em tempo logarítmico em relação ao tamanho do texto.

Trie

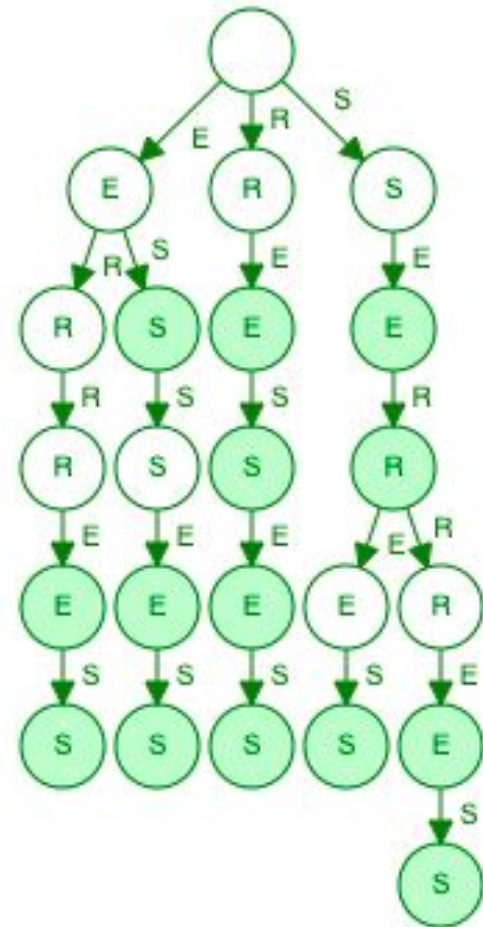
Patrícia

Pesquisa Digital - Trie

Exemplo:

$$\Sigma = \{e, r, s\}$$

S = { erre, erres, es, esse, essés, se, ser, serre, re, res, rese, reses, seres, serres }



<https://cmps-people.ok.ubc.ca/ylocet/DS/Trie.html>

Pesquisa Digital - Trie

Operações e Análise

– **Pesquisa** (Ex.: Existe a chave “serre”?)

- Seja k o tamanho da chave, então :

$$O(k(\log m + 1)) \xrightarrow{m=1} = O(k)$$

– **Inserção**

- Seja $k = k_1 + k_2$, então :

$$O(k_1 \log m + k_2 m)$$

- Complexidade independe de n
- Chaves de tamanho arbitrário e variável
- Possibilita a pesquisa de prefixos
- Aplicação direta inviável devido ao gasto com espaço
- Número excessivo de “ziguezagues” faz com que a pesquisa seja ineficiente

Pesquisa Digital - Trie

Observações

- Uma trie pode ser vista como um autômato finito determinístico, embora o símbolo em cada aresta seja frequentemente implícito na ordem dos galhos.
- Busca é mais rápida que em uma Árvore Binária
- Uma árvore Trie requer menos espaço quando contém um grande número de cadeias curtas, porque as chaves não são armazenadas de forma explícita e os nós das chaves iniciais comuns são compartilhados.

Pesquisa Digital - Patrícia

Origens

Practical Algorithm To Retrieve Information Coded In Alphanumeric
(Morrison/68)

A motivação inicial de Morrison foi otimizar a busca de arquivos em bibliotecas.

Características

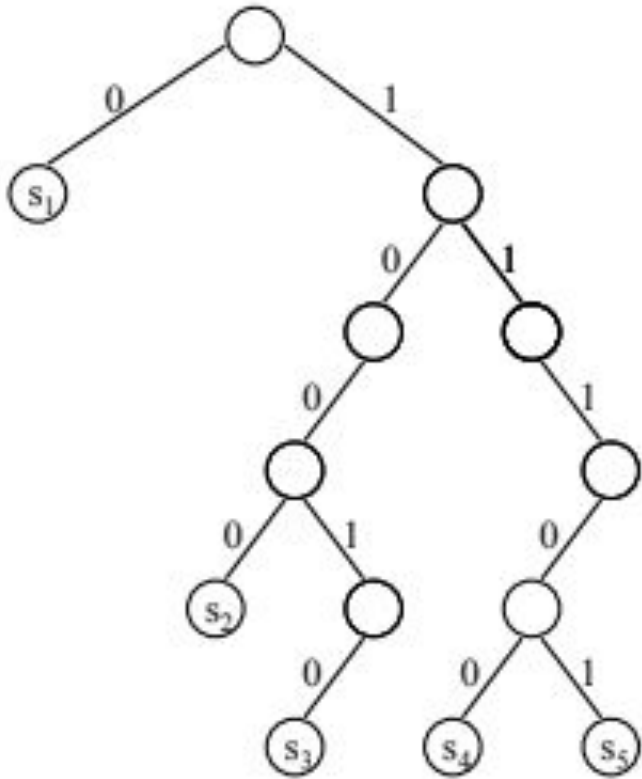
- Árvore binária de prefixo
- Não possui “ziguezagues”
- Cada nó possui um rótulo sobre qual dígito deve ser comparado

Pesquisa Digital - Patrícia

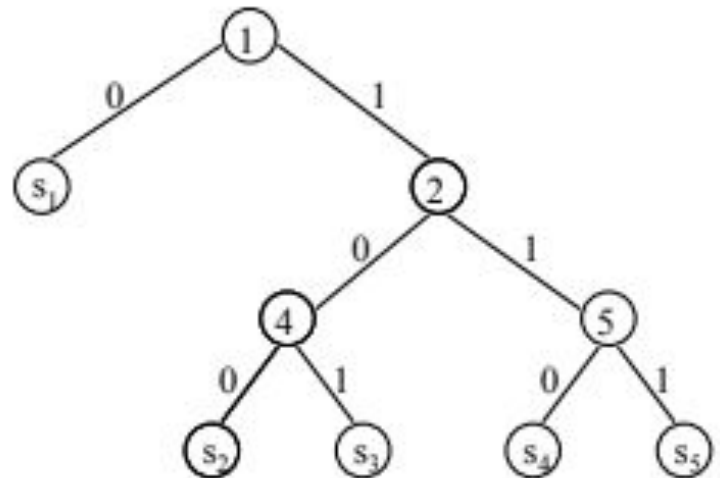
Exemplo $\Sigma = \{ 0,1 \}$ $S = \{ 0, 1000, 10010, 11100, 11101 \}$

S1 S2 S3 S4 S5

TRIE



PATRICIA

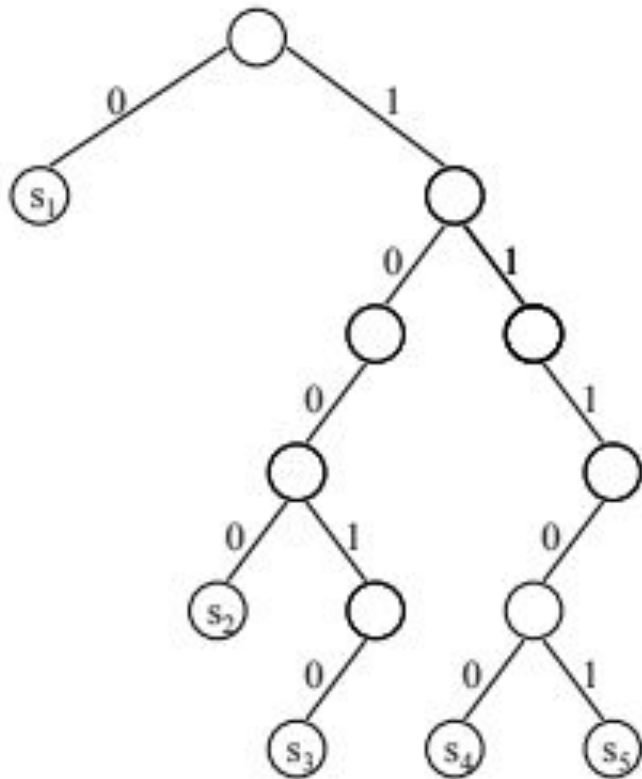


Pesquisa Digital - Patrícia

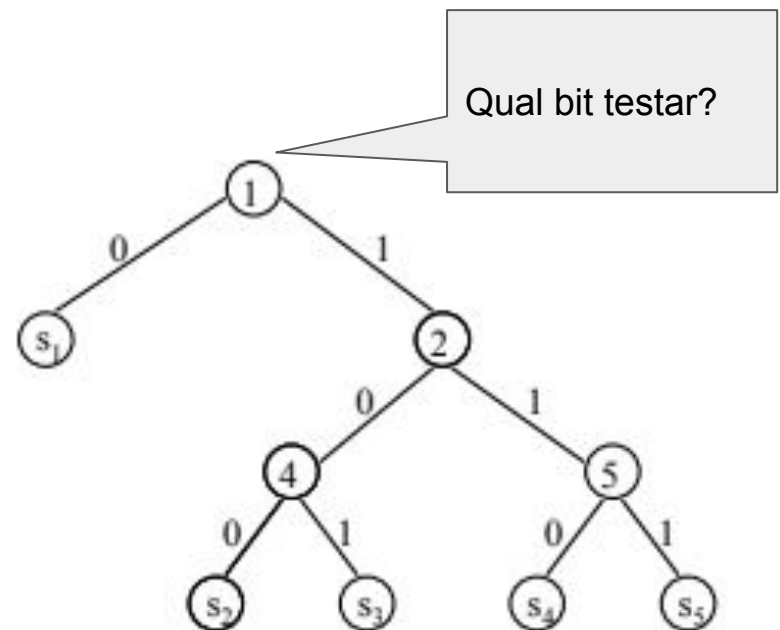
Exemplo $\Sigma = \{ 0, 1 \}$ $S = \{ 0, 1000, 10010, 11100, 11101 \}$

S1 S2 S3 S4 S5

TRIE



PATRICIA

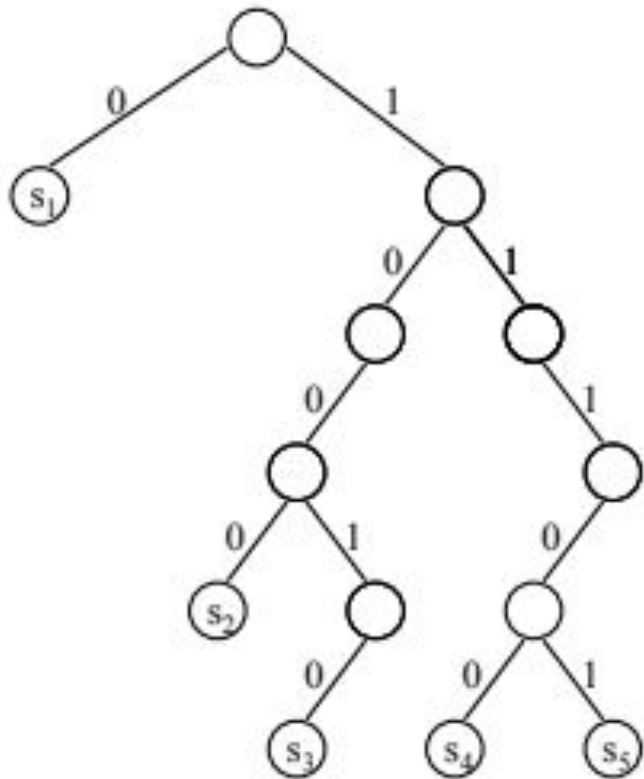


Pesquisa Digital - Patrícia

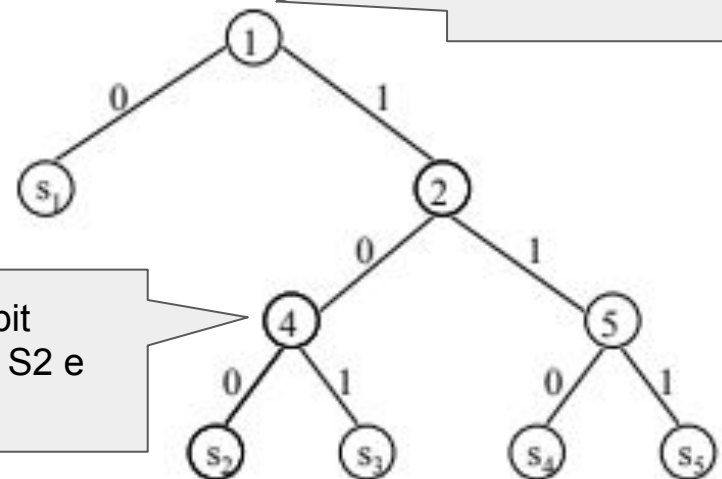
Exemplo $\Sigma = \{ 0, 1 \}$ $S = \{ 0, 1000, 10010, 11100, 11101 \}$

S1 S2 S3 S4 S5

TRIE



PATRICIA



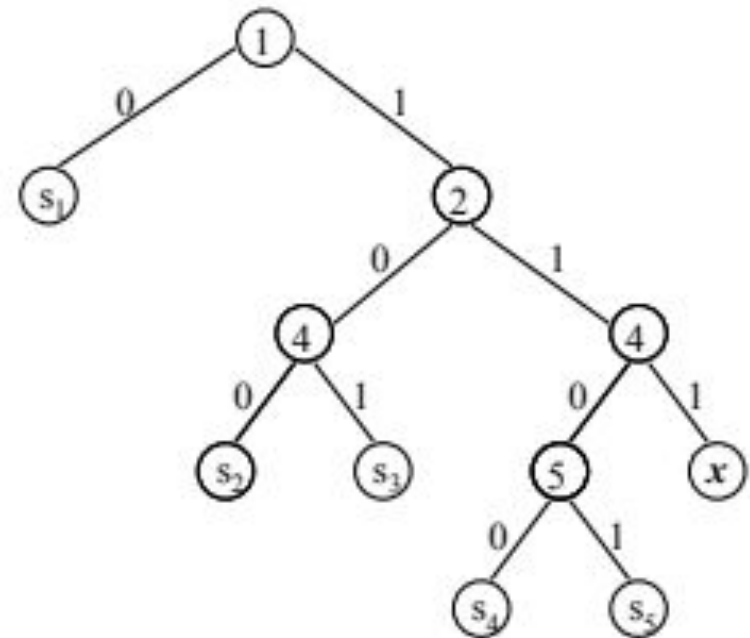
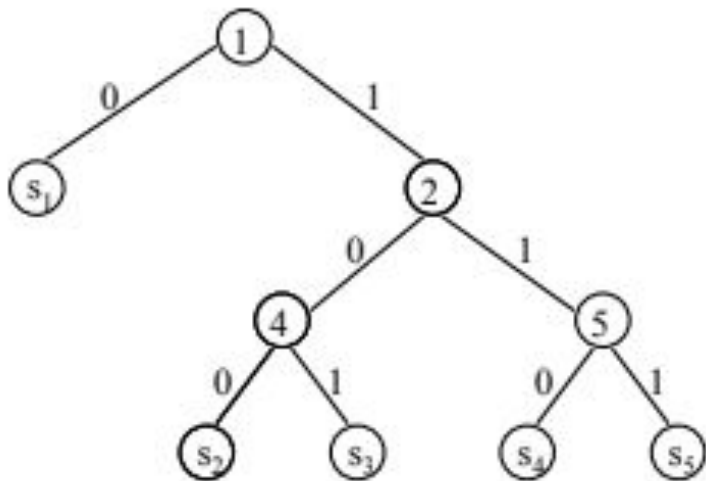
Qual bit testar?

O quarto bit
diferencia S2 e
S3

Pesquisa Digital - Patrícia

Exemplo $\Sigma = \{ 0, 1 \}$ $S = \{ 0, 1000, 10010, 11100, 11101 \}$
S1 S2 S3 S4 S5

Incluindo 11101



Pesquisa Digital - Patrícia

Pesquisa

Necessidade de
comparação entre
a chave e a folha

$O(K_{\max})$

Inserção

Inserir $x = 111101$

Pesquisa Digital - Patrícia

Observações

- Ao contrário da Trie, a árvore Patrícia não contém nós com apenas um filho.
- Cada nó tem dois filhos ou é um nó folha.
- O número de nós internos (não-folha) não ultrapassa o número de nós folhas.
- O tamanho de uma árvore Patrícia não depende do tamanho das chaves inseridas nela. Cada nova chave adiciona no máximo um novo nó na árvore.
- Diferente da árvore-B, a árvore Patrícia cresce lentamente mesmo com a inserção de um número grande de strings, por causa da alta compressão inerente da estrutura

Introdução



PUC Minas

Casamento de Padrões

O Casamento de padrões é uma busca por uma cadeia de símbolos em uma sequência maior

Ex.: busca de palavras em textos

Casamento de Padrões

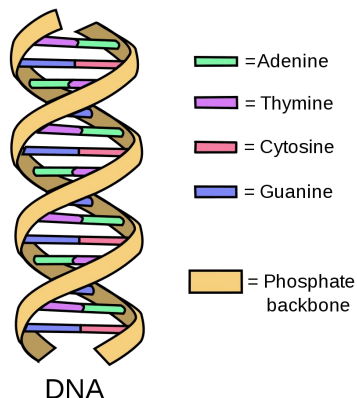
Os algoritmos são identificados como:

- Pattern searching
- String searching
- String matching

Casamento de Padrões

O Casamento de padrões não se limita a busca de texto.

Exemplos busca de DNA, detecção de intrusão, correção ortográfica, detecção de plágio, Casamento de voz, etc.



Casamento de Padrões

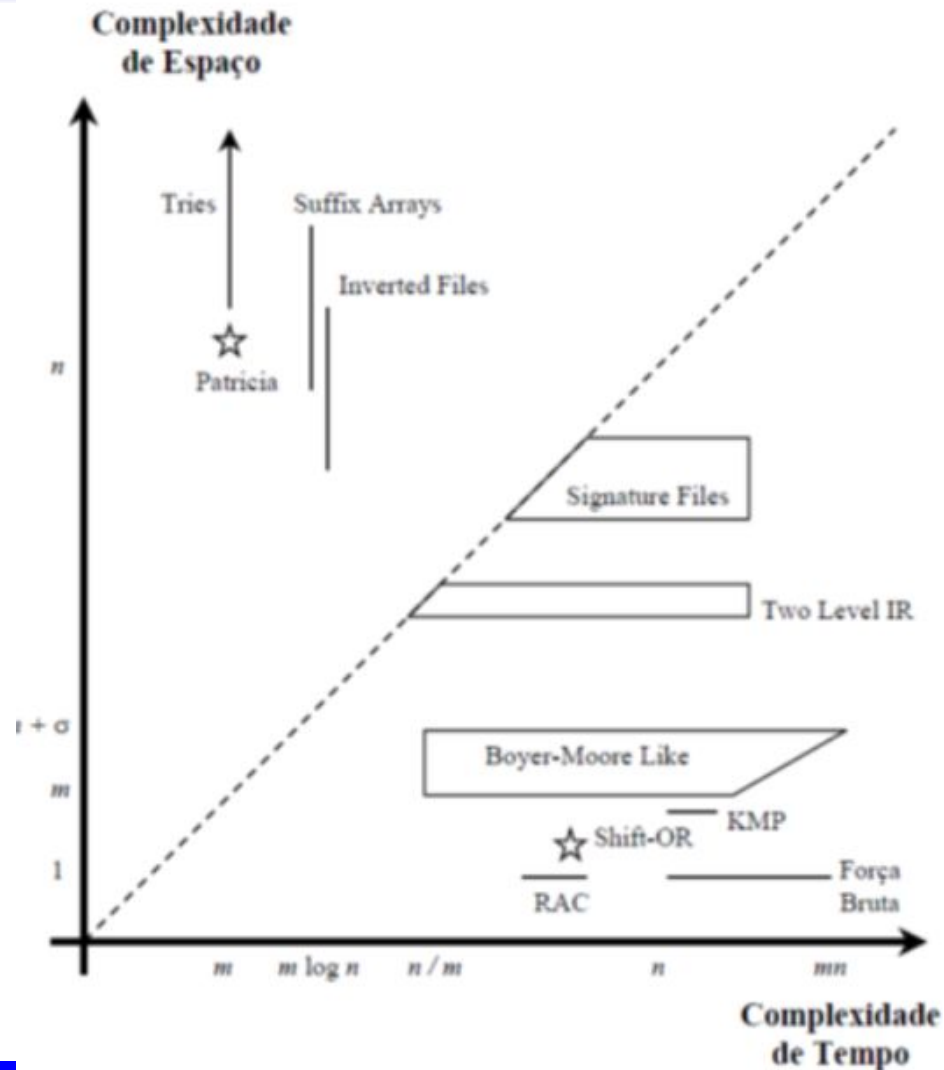
Casamento exato de padrões

- Força Bruta
- KMP
- Aho Corasick
- Boyer Moore

Casamento aproximado de padrões

- Distância de Levenshtein

Casamento de Padrões



Força Bruta



PUC Minas

Força Bruta

Motivação

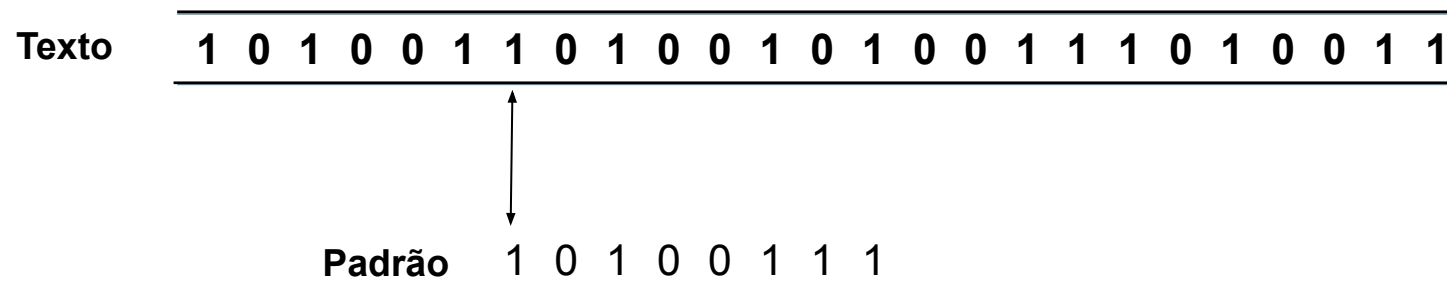
Texto

1 0 1 0 0 1 1 0 1 0 0 1 0 1 0 0 1 1 1 0 1 0 0 1 1

Padrão 1 0 1 0 0 1 1 1

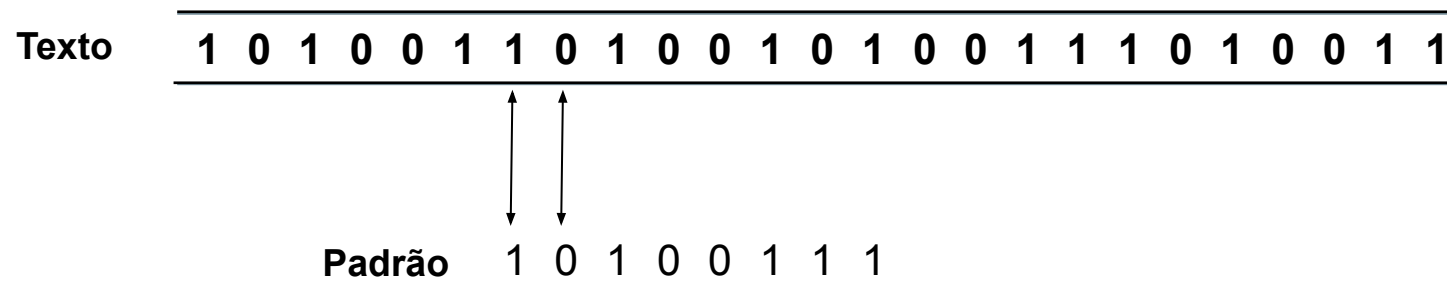
Força Bruta

- Motivação



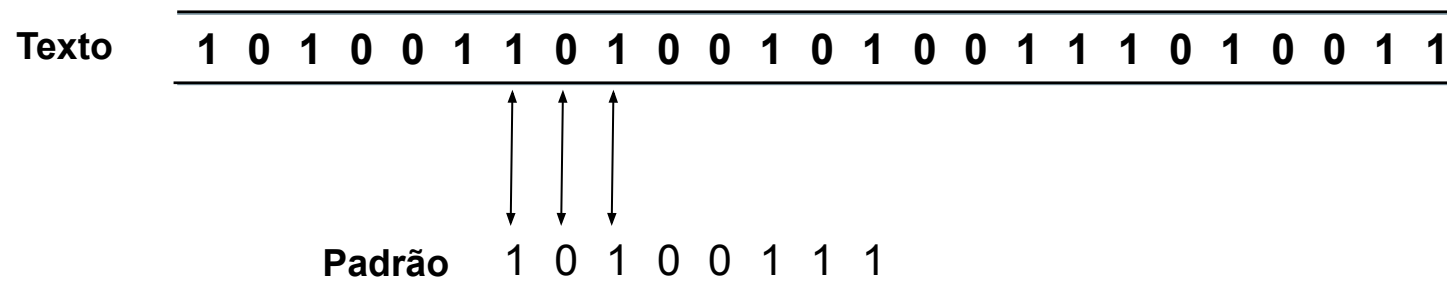
Força Bruta

- Motivação



Força Bruta

- Motivação



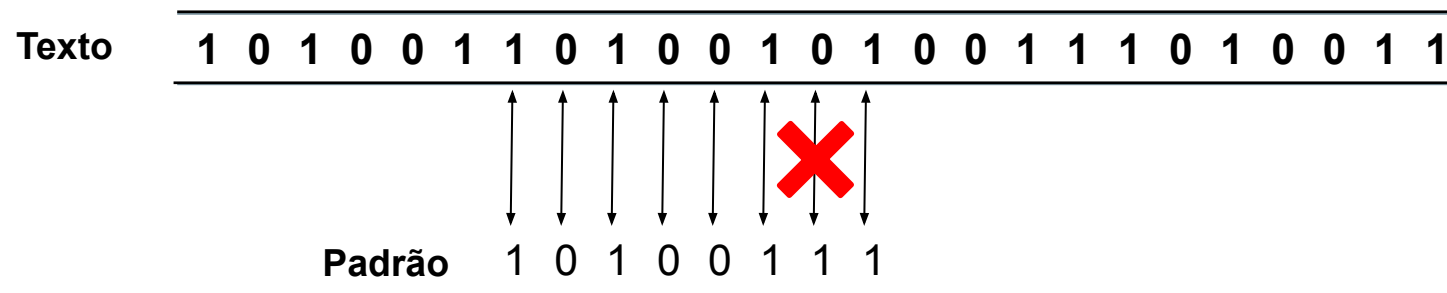
Força Bruta

- Motivação

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
							↑	↑	↑	↑	↑	↑	↑	↑											
Padrão							1	0	1	0	0	1	1	1											

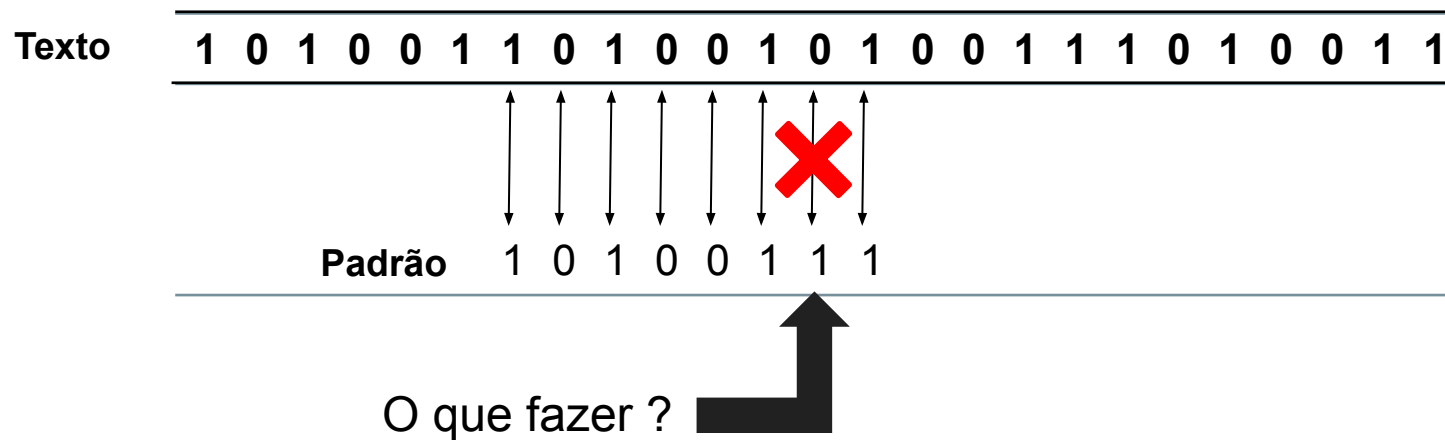
Força Bruta

- Motivação



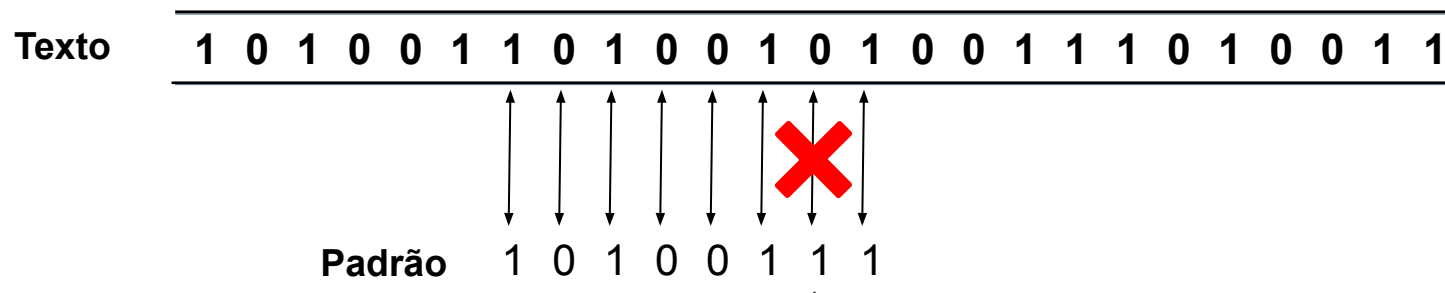
Força Bruta

- Motivação



Força Bruta

- Motivação



O que fazer ?



RESPOSTA:

- (i) Voltar ao início;
- (ii) Avançar uma posição; e
- (iii) Recomeçar

Força Bruta – Exemplo

Texto

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1																				
						1																			
							1																		
								1																	
									1																
										1															
											1														
												1													
													1												
														1											
															1										
																1									
																	1								
																		1							
																			1						
																				1					
																					1				
																						1			
																							1		
																								1	
																									1

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1																			
							1																		
								1																	
									1																
										1															
											1														
												1													
													1												
														1											
															1										
																1									
																	1								
																		1							
																			1						
																				1					
																					1				
																						1			
																							1		
																								1	
																									1

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1																	
									1																
										1															
											1														
												1													
													1												
														1											
															1										
																1									
																	1								
																		1							
																			1						
																				1					
																					1				
																						1			
																							1		
																								1	
																									1

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1													
													1												
														1											
															1										
																1									
																	1								
																		1							
																			1						
																				1					
																					1				
																						1			
																							1		
																								1	
																									1

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1													
													1												
														1											
															1										
																1									
																	1								
																		1							
																			1						
																				1					
																					1				
																						1			
																							1		
																								1	
																									1

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

[illegible]

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1													
													1												
														1											

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1	0												

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1	0	1												

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1	0	1	0										

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1	0	1	0	0									

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1	0	1	0	0	1								

Força Bruta – Exemplo

Texto
Padrão

1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
1	0	1	0	0	1	1	1																	
1																								
	1	0	1																					
		1	0	1																				
			1																					
				1																				
					1	0																		
						1	0	1	0	0	1	1												
							1																	
								1	0	1														
									1															
										1														
											1													
												1	0	1	0	0	1	1						

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																	
	1																								
		1	0	1																					
			1																						
				1																					
					1	0																			
						1	0	1	0	0	1	1													
							1																		
								1	0	1															
									1																
										1															
											1														
												1	0	1	0	0	1	1	1						

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1	1
Padrão	1	0	1	0	0	1	1	1																
	1																							
		1	0	1																				
			1																					
				1																				
					1	0																		
						1	0	1	0	0	1	1												
							1																	
								1	0	1														
									1															
										1														
											1													
												1	0	1	0	0	1	1	1					

Força Bruta – Exemplo

Texto	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1	1	# Comp.
Padrão	1	0	1	0	0	1	1	1																	8
	1																								1
		1	0	1																					3
			1																						1
				1																					1
					1	0																			2
						1	0	1	0	0	1	1													7
							1																		1
								1	0	1															3
									1																1
										1															1
											1	0	1	0	0	1	1	1							8
																									37

Força Bruta – Pior Caso

[illegible]

Força Bruta – Pior Caso

[illegible]

Força Bruta – Pior Caso

[illegible]

Força Bruta – Pior Caso

Texto	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Padrão	0	0	0	0	0	0	0	1																		
		0	0	0	0	0	0	0	1																	
			0	0	0	0	0	0	0	1																
				0	0	0	0	0	0	0	1															
					0	0	0	0	0	0	0	1														
						0	0	0	0	0	0	0	1													
							0	0	0	0	0	0	0	1												
								0	0	0	0	0	0	0	1											
									0	0	0	0	0	0	0	1										
										0	0	0	0	0	0	0	1									
											0	0	0	0	0	0	0	1								
												0	0	0	0	0	0	0	1							
													0	0	0	0	0	0	0	1						
														0	0	0	0	0	0	0	1					
																0	0	0	0	0	0	1				
																		0	0	0	0	0	0	1		
																					0	0	0	0	0	1

Força Bruta – Pior Caso

[illegible]

Força Bruta – Algoritmo

Reconhecimento de Padrões - Força Bruta

Seja : $T [0 .. n - 1] \rightarrow \text{texto}$

$P [0 .. m - 1] \rightarrow \text{padrão}$

ForçaBruta(T, P)

```
 $N \leftarrow \text{Tamanho}(T); \quad M \leftarrow \text{Tamanho}(P);$   
for ( $i = 0, j = 0; i < N - M \ \&\& \ j < M; i++, j++$ )  
    if ( $T[i] \neq P[j]$ ) {  $i = i - (j - 1); j = -1; \}$   
if ( $j == M$ ) return  $i - M$ ;  
else return  $i$ ;
```


Força Bruta – Resumo

Características

Não há fase de pré-processamento

Sempre avança apenas uma posição

Comparações podem ser feitas em qualquer ordem

Complexidade é $O(mn)$, m = tam. padrão, n = tam. texto

São esperadas $2n$ comparações

Análise

Melhor caso : $O(m)$

Pior caso : $O(mn)$

Caso médio : $O(n)$

⇒ Pior caso depende no número de “false starts”