

# Algoritmos e Estruturas de Dados III

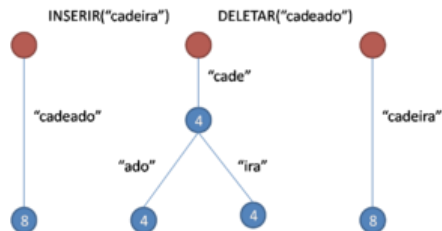
Aula 10.1 - Casamento de Padrões - Introdução e Força Bruta

Prof. Felipe Soares  
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)

# Casamento de Padrões

Árvore Digital (Trie - Fredkin/60)

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

Seja  $\Sigma$  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  $d_j$  de  $\Sigma$   $1 \leq j \leq m$
- 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 têm tamanho variável
- Localização de todas as ocorrências de determinada cadeia em um texto em tempo logarítmico em relação ao tamanho do texto.

Trie  
Patrícia

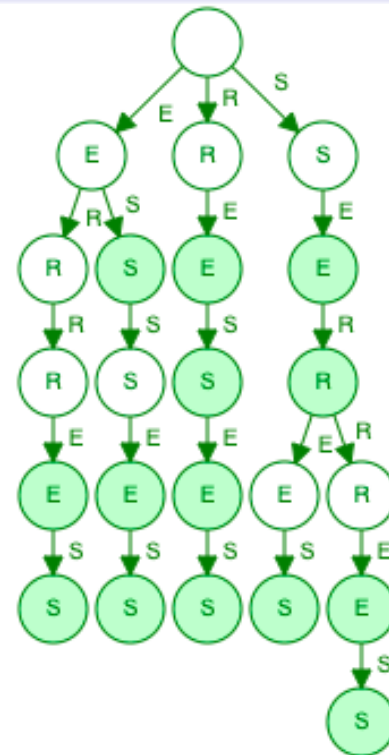


# Pesquisa Digital - Trie

## 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/ylucet/DS/Trie.html>

# Pesquisa Digital - Trie

## Operações e Análise

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

- Seja  $k$  o tamanho da chave,  $m$  o total de dígitos do alfabeto então :

$$O(k(\log m + 1))$$

– **Inserção**

- Seja  $k = k_1 + k_2$ , então :

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

- Complexidade independe de  $n$  (número de chaves)
- Chaves de tamanho arbitrário e variável
- Possibilita a pesquisa de prefixos



# 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

# Pesquisa Digital - Patrícia

## Origens

**P**ractical **A**lgorithm **T**o **R**etrieve Information **C**oded In **A**lphanumeric (Morrison/68)

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

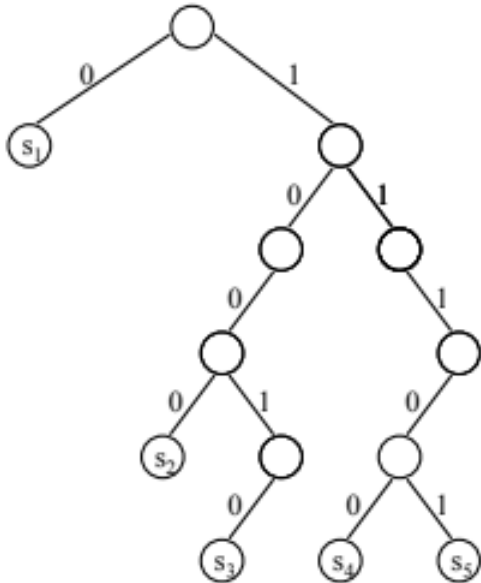
## Características

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

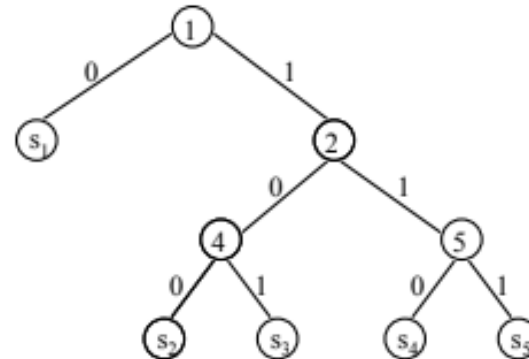
# Pesquisa Digital - Patrícia

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

S1 S2 S3 S4 S5



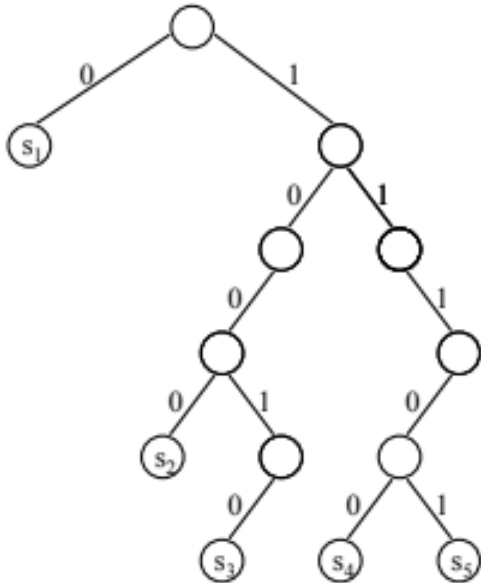
PATRICIA



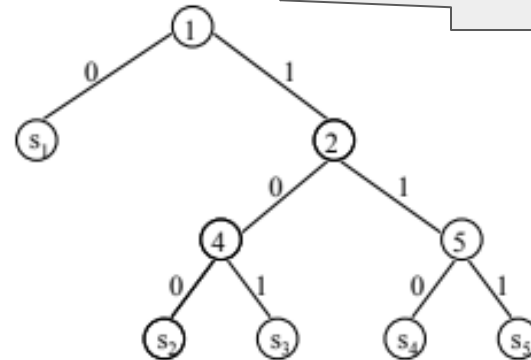
# Pesquisa Digital - Patrícia

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

S1 S2 S3 S4 S5



PATRICIA

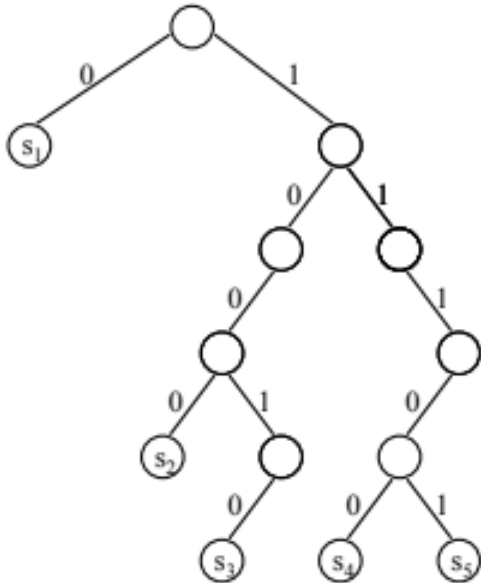


Qual bit testar?

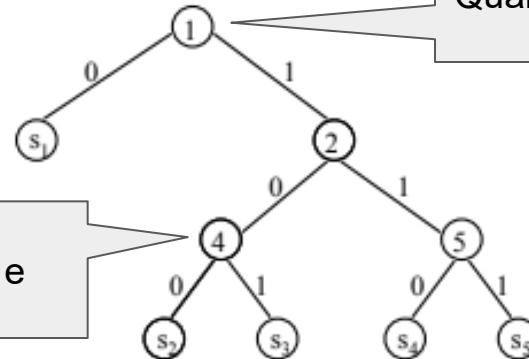
# Pesquisa Digital - Patrícia

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

**S1   S2   S3   S4   S5**



**PATRICIA**



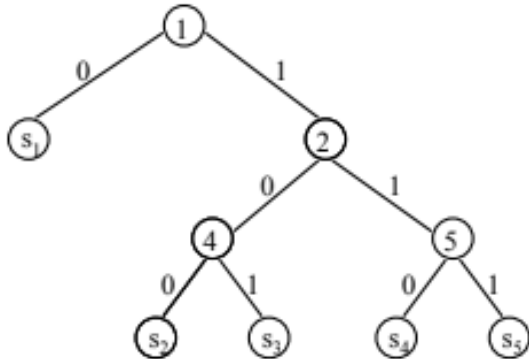
Qual bit testar?

O quarto bit  
diferencia S2 e  
S3

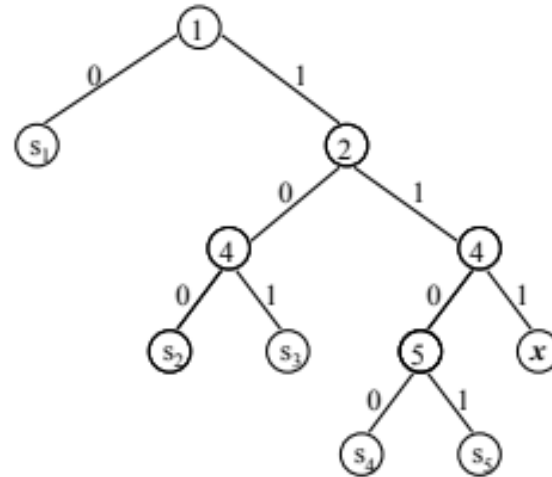
# Pesquisa Digital - Patrícia

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

**S1   S2   S3   S4   S5**



**Incluindo 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

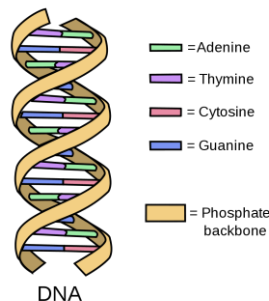
O Casamento de padrões é uma busca por uma cadeia de símbolos em uma sequência maior

Ex.: busca de palavras em textos

# Casamento de Padrões

O casamento de padrões não se limita à 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

# Força Bruta



**PUC Minas**

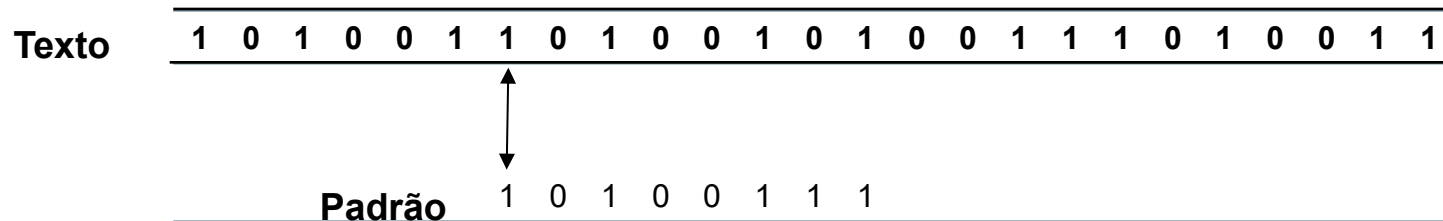
# Força Bruta

- 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																	

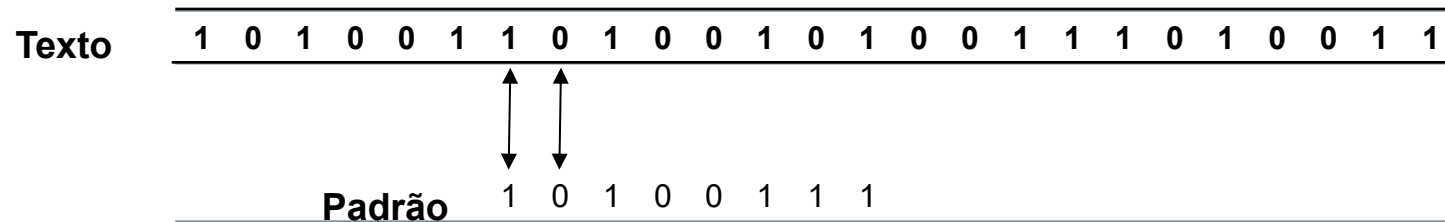
# Força Bruta – Algoritmo

- Motivação



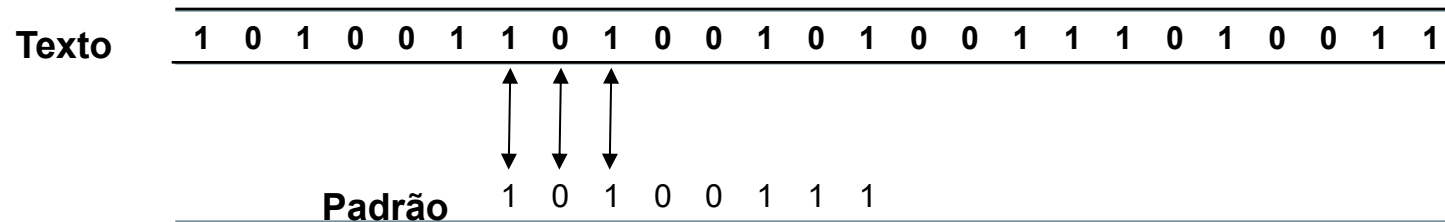
# Força Bruta – Algoritmo

- Motivação



# Força Bruta – Algoritmo

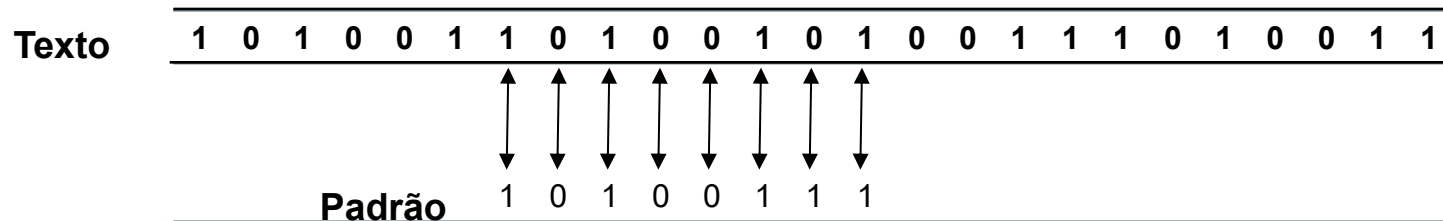
- Motivação





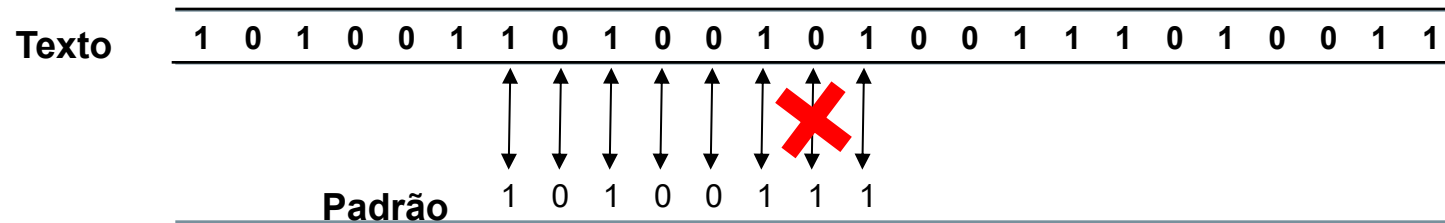
# Força Bruta – Algoritmo

- Motivação



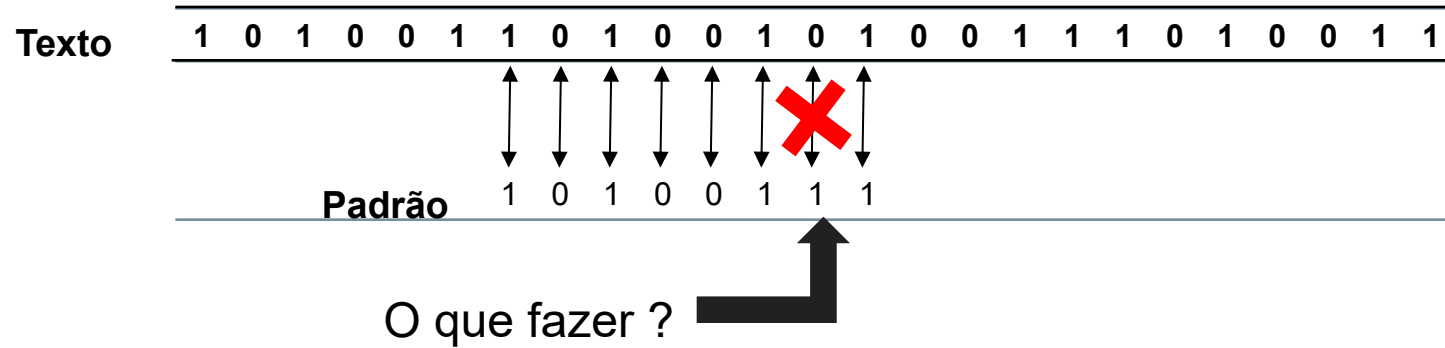
# Força Bruta – Algoritmo

- Motivação



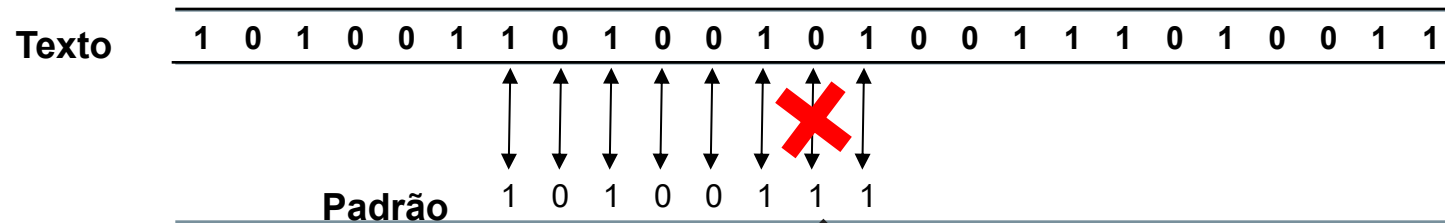
# Força Bruta – Algoritmo

- Motivação



# Força Bruta – Algoritmo

- Motivação



O que fazer ?

RESPOSTA:

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

# Força Bruta – Algoritmo

## Texto

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1



# Força Bruta – Algoritmo

**Texto  
Padrão**

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

**Texto  
Padrão**

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

**Texto  
Padrão**

[illegible]

Padrão: 1 0 1 0 0 1 1 1



# Força Bruta – Algoritmo

Diagram illustrating a memory space (256 bits) represented by a 25x25 grid. The top row is labeled "Texto Padrão" and contains a sequence of 25 bits: 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. The second row contains a sequence of 10 bits: 1, 0, 1, 0, 0, 1, 1, 1, followed by empty cells. The grid is used to illustrate the search for a specific pattern (10100111) within the memory space.

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1



# Força Bruta – Algoritmo

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																						
				1																					
					1																				
						1																			
							1																		
								1																	
									1																
										1															
											1														
												1													
													1												
														1											
															1										
																1									
																	1								
																		1							
																			1						
																				1					
																					1				
																						1			
																							1		
																								1	
																									1

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Diagrama de uma rede neural simples com 24 neurônios de entrada e 2 neurônios de saída. A primeira linha de pesos mostra a conexão entre os 24 neurônios de entrada e o primeiro neurônio de saída. Os pesos são: 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. A segunda linha mostra a conexão entre os 24 neurônios de entrada e o segundo neurônio de saída. Os pesos são: 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0. Os neurônios de saída são rotulados como "Texto" e "Padrão". Abaixo da grade, há uma barra de pesos para o padrão: Padrão: 1 0 1 0 0 1 1 1.

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Diagrama de uma rede neural simples com 24 neurônios na camada de entrada e 1 neurônio na camada de saída. A matriz de pesos ( $W$ ) é mostrada na parte superior, com valores binários (0 ou 1) em uma grade azul. A matriz de vieses ( $b$ ) está na parte inferior esquerda, também com valores binários. A saída ( $y$ ) é mostrada na parte inferior direita, com o valor 1 em um fundo vermelho.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
1	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
2	1	0	1	0	0	1	1	1																	
3		1																							
4			1	0	1																				
5				1																					
6					1																				
7						1	0																		
8							1	0	1																
9																									
10																									
11																									
12																									
13																									
14																									
15																									
16																									
17																									
18																									
19																									
20																									
21																									
22																									
23																									
24																									

Matriz de Vieses ( $b$ ):

1	1	0	1	0	0	1	1	1
---	---	---	---	---	---	---	---	---

Matriz de Saída ( $y$ ):

1	1
---	---

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1



# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Diagrama de uma rede neural simples com 20 neurônios de entrada e 20 neurônios de saída. A matriz de pesos é mostrada na parte superior, com alguns valores em vermelho. Abaixo, a matriz de ativação é mostrada, com alguns valores em vermelho. A saída final é mostrada na parte inferior.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0
2	1	0	1	0	0	1	1	1												
3		1																		
4			1	0	1															
5			1																	
6				1																
7					1	0														
8						1	0	1	0	0	1	1								
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Texto  
Padrão

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão



# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Texto  
Padrão

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Texto  
Padrão

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

Texto  
Padrão

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

[illegible]

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

## Texto Padrão

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1



# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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																					
				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					

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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																					
				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					

Padrão: 1 0 1 0 0 1 1 1

# Força Bruta – Algoritmo

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

Padrão: 1 0 1 0 0 1 1 1

# Comp.

8  
1  
3  
1  
1  
2  
7  
1  
3  
1  
1  
8

---

37

# Força Bruta – Pior Caso

[illegible]



# Força Bruta – Pior Caso

[illegible]

# Força Bruta – Pior Caso

[illegible]

# Força Bruta – Pior Caso

[illegible]

# Força Bruta – Pior Caso

Texto  
Padrão

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

# Comp.

8  
8  
8  
8  
8  
8  
8  
8  
8  
8  
8  
8  
8  
8

96

# Força Bruta – Algoritmo

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”