

# ESTRUTURAS DE DADOS II

---

MSC. DANIELE CARVALHO OLIVEIRA

MESTRE EM CIÊNCIA DA COMPUTAÇÃO – UFU

BACHAREL EM CIÊNCIA DA COMPUTAÇÃO - UFJF

# 2

## BUSCA EM TEXTO

---



# PROBLEMA

---

- Problema da Pesquisa em Strings
  - encontrar todas as ocorrências de um padrão de comprimento  $m$  num texto de comprimento  $n$
  - texto e padrão são cadeias de caracteres sobre o mesmo alfabeto  $\Sigma$
  - com  $m$  fixo, esforço computacional é  $O(n)$  no pior caso
  - é preciso inspeccionar  $n-m+1$  caracteres no pior caso
  - factor constante na eficiência depende muito do algoritmo

# ALGORITMO EXAUSTIVO

- Princípio: tentar identificar o padrão em cada subcadeia de comprimento  $m$  do texto
- Exemplo

1	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	1	0	0	0					
1	0	1	0	0	1	1	1																								
	1	0	1	0	0	1	1	1																							
		1	0	1	0	0	1	1	1																						
			1	0	1	0	0	1	1	1																					
				1	0	1	0	0	1	1	1																				
					1	0	1	0	0	1	1	1																			
						1	0	1	0	0	1	1	1																		
							1	0	1	0	0	1	1	1																	
...																															
																					1	0	1	0	0	1	1	1			



## 5 NAIVE-STRING-MATCHER

---

Naive-String-Matcher(T,P)

1.  $n = T.\text{comprimento}$
2.  $m = P.\text{comprimento}$
3. For  $s = 0$  to  $n-m$
4.     if  $P[1 .. m] == T[s + 1 .. s + m]$
5.         imprimir “Padrão ocorre com deslocamento”  $s$

## 6 EXERCÍCIO

---

- Implemente o algoritmo Naive String Matcher
- Altere o algoritmo para aceitar caractere coringa no Padrão
  - O caractere coringa pode ocorrer várias vezes no Padrão, nenhuma vez no Texto

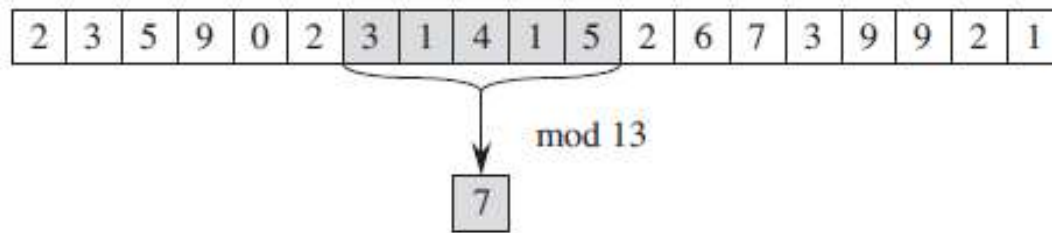
# 7 ALGORITMO RABIN-KARP

---

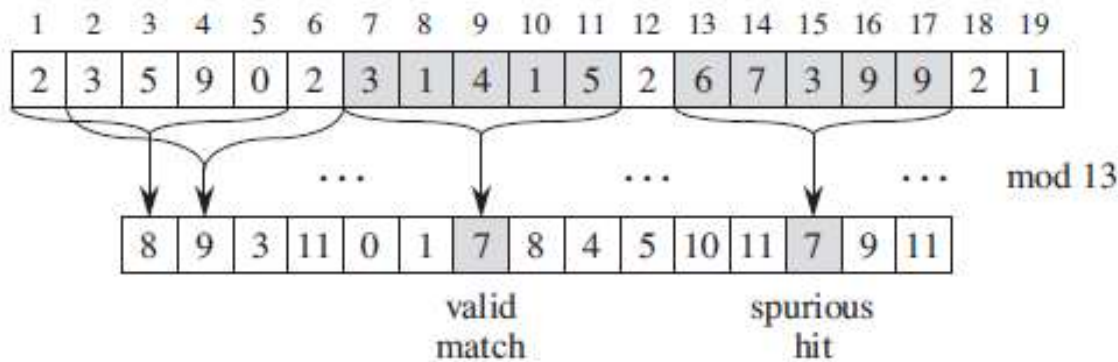




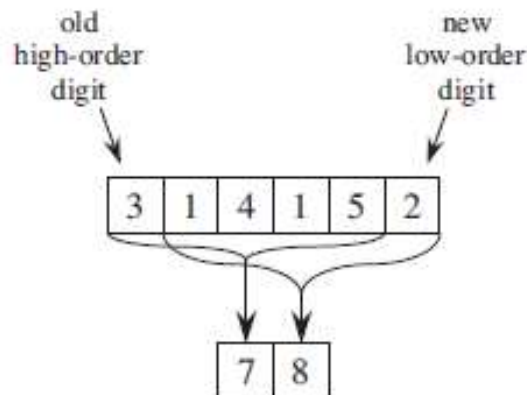
8



(a)



(b)



$$\begin{aligned}
 14152 &\equiv (31415 - 3 \cdot 10000) \cdot 10 + 2 \pmod{13} \\
 &\equiv (7 - 3 \cdot 3) \cdot 10 + 2 \pmod{13} \\
 &\equiv 8 \pmod{13}
 \end{aligned}$$



## 9 RABIN KARP MATCHER

*Rabin – Karp – Matcher( $T, P, d, q$ )*

1.  $n = T.comprimento$
2.  $m = P.comprimento$
3.  $h = d^{m-1} \bmod q$
4.  $p = 0$
5.  $t_0 = 0$
6. *for*  $i = 1$  *to*  $m$
7.      $p = (dp + P[i]) \bmod q$
8.      $t_0 = (dt_0 + T[i]) \bmod q$
9. *for*  $s = 0$  *to*  $n - m$
10.     *if*  $p == t_s$
11.         *if*  $P[1 .. m] = T[s + 1 .. s + m]$
12.             *imprima* "Padrão ocorre com deslocamento"  $s$
13.     *if*  $s < n - m$
14.          $t_{s+1} = (d(t_s - T[s + 1]h) + T[s + m + 1]) \bmod q$



# Trabalho

- Implemente o algoritmo Rabin Karp Matcher
- Estenda o algoritmo para dado um texto, encontrar a ocorrência de qualquer padrão dentre um conjunto de padrões de comprimentos diferentes.



# FIM DA AULA 5

---

Próxima aula:  
Notação  $O$