

Algoritmos e Estruturas de Dados II

String matching

Prof. Flávio José M. Coelho

fcoelho@uea.edu.br

Objetivos

1. Entender o que é **String Matching**


Entender os algoritmos:

2. “ingênuo” para string matching

String Matching

String Matching

“pattern”



Text-editing programs frequently need to find all occurrences of a **pattern** in the text. Typically, the text is a document being edited, and the **pattern** searched for is a particular word supplied by the user. Efficient algorithms for this problem—called “string matching”—can greatly aid the responsiveness of the text-editing program. Among their many other applications, string-matching algorithms search for particular **patterns** in DNA sequences.

We formalize the string-matching problem as follows. We assume that the text is an array of length n and that the **pattern** is an array of length m . We further assume that the elements of P and T are characters drawn from a finite alphabet.

String Matching

Suponha que um vetor $T[1..n]$ de comprimento n (texto), e vetor $P[1..m]$ (padrão) de comprimento $m \leq n$.

String Matching

Suponha que um vetor $T[1..n]$ de comprimento n (texto), e vetor $P[1..m]$ (padrão) de comprimento $m \leq n$.

P e T são **strings de caracteres** geradas a partir de um alfabeto finito Σ .

String Matching

Dizemos que P **ocorre com deslocamento** s em $T[1..n]$, se $0 \leq s \leq n - m$ e $T[s+1..s+m] = P[1..m]$.

texto T

a	b	c	a	b	a	a	b	c	a	b	a	c
---	---	---	---	---	---	---	---	---	---	---	---	---

padrão $P \xrightarrow{s=3}$

a	b	a	a
---	---	---	---

String Matching

Se P ocorre com um deslocamento s em T , então s é um **deslocamento válido** (*valid shift*);

String Matching

Se P ocorre com um deslocamento s em T , então s é um **deslocamento válido** (*valid shift*); caso contrário, s é um **deslocamento inválido**.

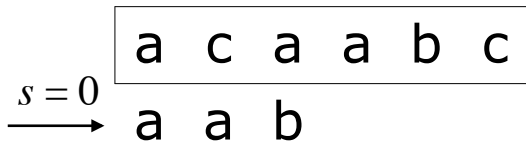
String Matching

Se P ocorre com um deslocamento s em T , então s é um **deslocamento válido** (*valid shift*); caso contrário, s é um **deslocamento inválido**.

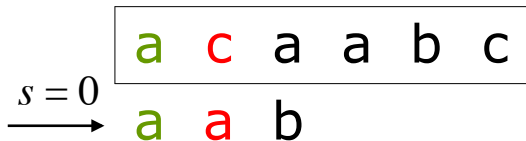
Em um problema de **String Matching** deseja-se achar todos os deslocamentos válidos com os quais um padrão P ocorre em um dado texto T .

Algoritmo “ingênuo”

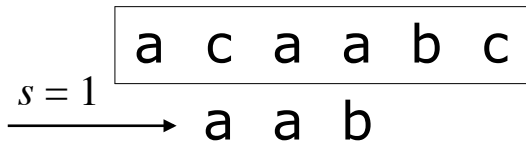
Algoritmo “ingênuo”



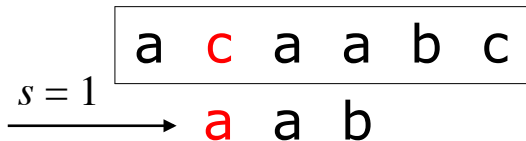
Algoritmo “ingênuo”



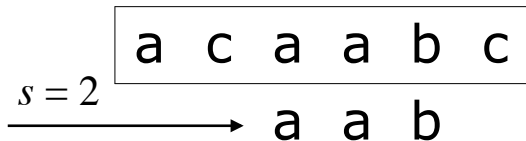
Algoritmo “ingênuo”



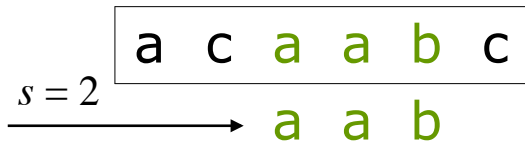
Algoritmo “ingênuo”



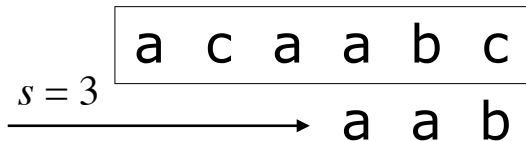
Algoritmo “ingênuo”



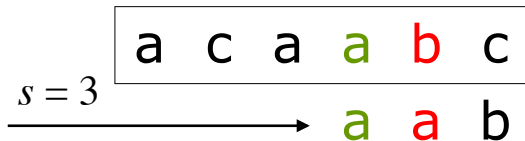
Algoritmo “ingênuo”



Algoritmo “ingênuo”



Algoritmo “ingênuo”



STRING-MATCHER-INGENUO(T, P)





```
1   $n = T.tamanho$   
2   $m = P.tamanho$   
3  para  $s = 0$  até  $n - m$   
4      se  $P[1..m] == T[s + 1..s + m]$   
5          imprima “Padrão ocorre  
           com deslocamento”  $s$ 
```

Algoritmo “ingênuo”

Complexidade: $O((n - m + 1)m)$

Não é um método eficiente em comparação com outros.

Referências

-  T. H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, 3rd edition, MIT Press, 2010
-  A. Levitin. Introduction to the Design and Analysis of Algorithms. 3rd edition. Addison-Wesley, 2007
-  R. Sedgewick, K. Wayne. Algorithms. 4th edition, Addison-Wesley Professional, 2011
-  N. Ziviani. Projeto de Algoritmos com Implementação em Pascal C. Cengage Learning, 2012

Onde obter este material:

`est.uea.edu.br/fcoelho`