

Casamento aproximado

- Algoritmos buscam reconhecer padrões parecidos, considerando:
 - Erros ortográficos (ex.: atensão)
 - Mesmos fonemas (ex.: paço e passo)
 - Variações ortográficas (ex.: Luiz e Luís)
 - Etc.

Casamento aproximado

- Aplicações:
 - Corretores ortográficos
 - Comparação de DNA
 - Filtragem de SPAM
 - OCR (optical character recognition)

Casamento aproximado

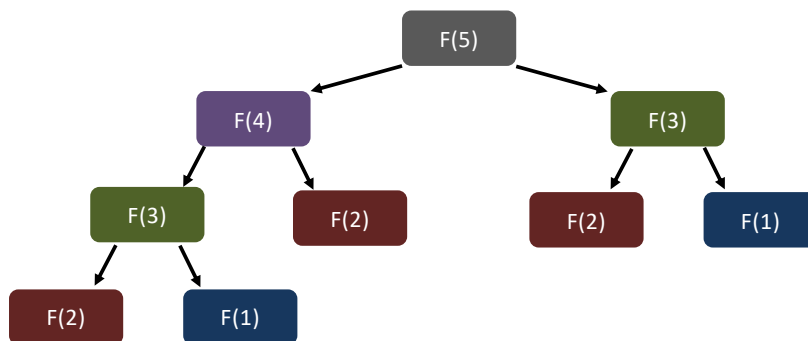
- Distância de edição
 - Mede a diferença entre duas sequências
 - A diferença é dada pelo número de edições (inserções, exclusões e substituições) necessárias para transformar uma sequência em outra.
- Exemplos:
 - Gato → Pato (1 edição)
 - Paço → Passo (2 edições)

Casamento aproximado

- Programação dinâmica
 - Método de solução de problemas por meio da sua decomposição em problemas menores
 - Requisitos:
 - Subestrutura ótima – a solução pode ser obtida pela combinação da solução dos subproblemas
 - Sobreposição de subproblemas – a solução do problema pode envolver resolver os mesmos subproblemas várias vezes

Casamento aproximado


- Programação dinâmica
 - Exemplo de problema
 - Fibonacci – $F(n) = F(n-1) + F(n-2)$



Casamento aproximado

- Programação dinâmica
 - Solução – calcular (e guardar em uma tabela) os valores dos “subproblemas”

F(1)	F(2)	F(3)	F(4)	F(5)
1	1	2	3	5



Casamento aproximado

- Distância de Levenshtein (1965)

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{se } \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + C \end{cases} & \text{nos outros casos} \end{cases}$$

em que $C = 0$, se $a_i = b_j$, ou $C = 1$, se $a_i \neq b_j$.

Casamento aproximado

- Exemplo 1:
CASACO x CASCAO

		C	A	S	A	C	O
	0	1	2	3	4	5	6
C	1						
A	2						
S	3						
C	4						
A	5						
O	6						

Casamento aproximado

- Exemplo 1:
CASACO x CASCAO

		C	A	S	A	C	O
C A S C A O	0	1	2	3	4	5	6
	1	0	1	2	3	4	5
	2	1	0	1	2	3	4
	3	2	1	0	1	2	3
	4	3	2	1	1	1	2
	5	4	3	2	2	2	2
	6	5	4	3	3	2	2

Casamento aproximado

- Exemplo 1:
PIGARRO x CIGANO

		P	I	G	A	R	R	O
C I G A N O	0	1	2	3	4	5	6	7
	1	1	2	3	4	5	6	7
	2	2	1	2	3	4	5	6
	3	3	2	1	2	3	4	5
	4	4	3	2	1	2	3	4
	5	5	4	3	2	2	3	4
	6	6	5	4	3	3	3	3