Análise de Algoritmos
Aula 14

Prof. Murilo V. G. da Silva

DINF/UFPR

(material da disciplina: André Guedes, Renato Carmo, Murilo da Silva)

Alinhamento de Sequências e Distância de Edição

Distância de Edição

Dadas duas palaras x e y, qual o mínimo de operações de edição (inserção, remoção e troca de letras) para transformar uma no outra?

- Seja $x = (x_1, \ldots, x_n)$ uma palavra sobre um alfabeto Σ

- Seja $x = (x_1, \ldots, x_n)$ uma palavra sobre um alfabeto Σ
- Seja $\star \notin \Sigma$

- Seja $x = (x_1, \ldots, x_n)$ uma palavra sobre um alfabeto Σ
- Seja $\star \notin \Sigma$

Expansão de uma palavra

Dizemos que a palavra $x' \in \Sigma \cup \{\star\}$ é uma *expansão* de x se a palavra obtida ao retirar todos os $\star$ de x' é x.

- Seja $x = (x_1, \ldots, x_n)$ uma palavra sobre um alfabeto Σ
- Seja $\star \notin \Sigma$

Expansão de uma palavra

Dizemos que a palavra $x' \in \Sigma \cup \{\star\}$ é uma *expansão* de x se a palavra obtida ao retirar todos os $\star$ de x' é x.

Exemplo:

- $\texttt{ca}\star\texttt{ra}\star$ é uma expansão de $\texttt{cara}$.

- Seja $x = (x_1, \ldots, x_n)$ uma palavra sobre um alfabeto Σ
- Seja $\star \notin \Sigma$

Expansão de uma palavra

Dizemos que a palavra $x' \in \Sigma \cup \{\star\}$ é uma *expansão* de x se a palavra obtida ao retirar todos os $\star$ de x' é x.

Exemplo:

- `ca⋆ra⋆` é uma expansão de `cara`.
- `cald⋆o` é uma expansão de `caldo`.

Alinhamento de palavras

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Alinhamento de Sequências e Distância de Edição

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Se

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Se

1. $|x'| = |y'|$, e

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Se

1. $|x'| = |y'|$, e
2. x' é expansão de x, e

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Se

1. $|x'| = |y'|$, e
2. x' é expansão de x, e
3. y' é expansão de y.

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Se

1. $|x'| = |y'|$, e
2. x' é expansão de x, e
3. y' é expansão de y.

- $(\texttt{ca}\star\texttt{ra}\star, \texttt{cald}\star\texttt{o})$ é um alinhamento de $(\texttt{cara}, \texttt{caldo})$.

Alinhamento de palavras

Dadas duas palavras x e y sobre Σ, dizemos que (x', y') é um *alinhamento* de (x, y)

Se

1. $|x'| = |y'|$, e
2. x' é expansão de x, e
3. y' é expansão de y.

- $(\texttt{ca}\star\texttt{ra}\star, \texttt{cald}\star\texttt{o})$ é um alinhamento de $(\texttt{cara}, \texttt{caldo})$.
- $(\texttt{ca}\star\texttt{ra}, \texttt{caldo})$ é outro alinhamento de $(\texttt{cara}, \texttt{caldo})$.

Dados $\sigma, \tau \in (\Sigma \cup \star)$ definimos

$$d(\sigma, \tau) = \begin{cases} 0, & \text{se } \sigma = \tau, \\ 1, & \text{se } \sigma \neq \tau \end{cases}$$

Dados $\sigma, \tau \in (\Sigma \cup \star)$ definimos

$$d(\sigma, \tau) = \begin{cases} 0, & \text{se } \sigma = \tau, \\ 1, & \text{se } \sigma \neq \tau \end{cases}$$

Distância de Hamming

A *Distância de Hamming* entre duas palavras x e y de mesmo tamanho é número de posições em que x e y diferem, isto é,

Dados $\sigma, \tau \in (\Sigma \cup \star)$ definimos

$$d(\sigma, \tau) = \begin{cases} 0, & \text{se } \sigma = \tau, \\ 1, & \text{se } \sigma \neq \tau \end{cases}$$

Distância de Hamming

A *Distância de Hamming* entre duas palavras x e y de mesmo tamanho é número de posições em que x e y diferem, isto é,

$$d_H(x, y) = \sum_{i=1}^{|x|} d(x_i, y_i).$$

Dados $\sigma, \tau \in (\Sigma \cup \star)$ definimos

$$d(\sigma,\tau) = \begin{cases} 0, & \text{se } \sigma = \tau, \\ 1, & \text{se } \sigma \neq \tau \end{cases}$$

Distância de Hamming

A *Distância de Hamming* entre duas palavras x e y de mesmo tamanho é número de posições em que x e y diferem, isto é,

$$d_H(x,y) = \sum_{i=1}^{|x|} d(x_i, y_i).$$

Exemplo:

Dados $\sigma, \tau \in (\Sigma \cup \star)$ definimos

$$d(\sigma,\tau) = \begin{cases} 0, & \text{se } \sigma = \tau, \\ 1, & \text{se } \sigma \neq \tau \end{cases}$$

Distância de Hamming

A *Distância de Hamming* entre duas palavras x e y de mesmo tamanho é número de posições em que x e y diferem, isto é,

$$d_H(x,y) = \sum_{i=1}^{|x|} d(x_i, y_i).$$

Exemplo:

$$d_H(\texttt{ca}\star\texttt{ra}\star, \texttt{cald}\star\texttt{o}) \quad = \quad 4$$

Dados $\sigma, \tau \in (\Sigma \cup \star)$ definimos

$$d(\sigma, \tau) = \begin{cases} 0, & \text{se } \sigma = \tau, \\ 1, & \text{se } \sigma \neq \tau \end{cases}$$

Distância de Hamming

A *Distância de Hamming* entre duas palavras x e y de mesmo tamanho é número de posições em que x e y diferem, isto é,

$$d_H(x, y) = \sum_{i=1}^{|x|} d(x_i, y_i).$$

Exemplo:

$$\begin{aligned} d_H(\texttt{ca}\star\texttt{ra}\star, \texttt{cald}\star\texttt{o}) &= 4 \\ d_H(\texttt{ca}\star\texttt{ra}, \texttt{caldo}) &= 3 \end{aligned}$$

Alinhamento de Sequências e Distância de Edição

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Alinhamento de Sequências e Distância de Edição

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Relação entre alinhamento e edição

Para cada $i \in [1..|x'|]$,

Um alinhamento (x', y') de (x, y) é um “roteiro de edição” para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Relação entre alinhamento e edição

Para cada $i \in [1..|x'|]$,

$x_i = \star$ e $y_i \neq \star$ indica a inserção de y_i entre x_{i-1} e x_i,

Alinhamento de Sequências e Distância de Edição

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Relação entre alinhamento e edição

Para cada $i \in [1..|x'|]$,

$x_i = \star$ e $y_i \neq \star$ indica a inserção de y_i entre x_{i-1} e x_i,

$x_i \neq \star$ e $y_i = \star$ indica a remoção de x_i,

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Relação entre alinhamento e edição

Para cada $i \in [1..|x'|]$,

$x_i = \star$ e $y_i \neq \star$ indica a inserção de y_i entre x_{i-1} e x_i,

$x_i \neq \star$ e $y_i = \star$ indica a remoção de x_i,

$x_i \neq \star$ e $y_i \neq \star$ e $x_i \neq y_i$ indica a troca de x_i por y_i.

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Relação entre alinhamento e edição

Para cada $i \in [1..|x'|]$,

$x_i = \star$ e $y_i \neq \star$ indica a inserção de y_i entre x_{i-1} e x_i,

$x_i \neq \star$ e $y_i = \star$ indica a remoção de x_i,

$x_i \neq \star$ e $y_i \neq \star$ e $x_i \neq y_i$ indica a troca de x_i por y_i.

- A distância de Hamming entre x' e y' é o número de operações necessárias para transformar x em y.

Um alinhamento (x', y') de (x, y) é um "roteiro de edição" para transformar x em y

- Roteiro esse usando operações de inserção, remoção e troca de letras.

Relação entre alinhamento e edição

Para cada $i \in [1..|x'|]$,

$x_i = \star$ e $y_i \neq \star$ indica a inserção de y_i entre x_{i-1} e x_i,

$x_i \neq \star$ e $y_i = \star$ indica a remoção de x_i,

$x_i \neq \star$ e $y_i \neq \star$ e $x_i \neq y_i$ indica a troca de x_i por y_i.

- A distância de Hamming entre x' e y' é o número de operações necessárias para transformar x em y.
- O alinhamento (x, y) com a menor distância de Hamming possível nos dá exatamente isso

Alinhamento de Sequências e Distância de Edição

Alinhamento ótimo: um alinhamento (x', y') de (x, y) é ótimo se $d_H(x', y')$ tem o menor valor possível.

Alinhamento ótimo: um alinhamento (x', y') de (x, y) é ótimo se $d_H(x', y')$ tem o menor valor possível.

(se existem vários possíveis, o ótimo deve minimizar $|x'|$)

Alinhamento ótimo: um alinhamento (x', y') de (x, y) é ótimo se $d_H(x', y')$ tem o menor valor possível.
(se existem vários possíveis, o ótimo deve minimizar $|x'|$)

Distância de Edição

A *Distância de Edição* (ou *Distância de Levenshtein*) entre duas palavras x e y é a distância de Hamming de um alinhamento ótimo de (x, y), isto é,

$$d_E(x, y) = \min\left\{d_H(x', y') \mid (x', y') \text{ é alinhamento de } (x, y)\right\}.$$

Alinhamento de Sequências e Distância de Edição

Alinhamento ótimo: um alinhamento (x', y') de (x, y) é ótimo se $d_H(x', y')$ tem o menor valor possível.

(se existem vários possíveis, o ótimo deve minimizar $|x'|$)

Distância de Edição

A *Distância de Edição* (ou *Distância de Levenshtein*) entre duas palavras x e y é a distância de Hamming de um alinhamento ótimo de (x, y), isto é,

$$d_E(x, y) = \min \{ d_H(x', y') \mid (x', y') \text{ é alinhamento de } (x, y) \}.$$

O Problema da Distância de Edição (DE)

Alinhamento de Sequências e Distância de Edição

Alinhamento ótimo: um alinhamento (x', y') de (x, y) é ótimo se $d_H(x', y')$ tem o menor valor possível.
(se existem vários possíveis, o ótimo deve minimizar $|x'|$)

Distância de Edição

A *Distância de Edição* (ou *Distância de Levenshtein*) entre duas palavras x e y é a distância de Hamming de um alinhamento ótimo de (x, y), isto é,

$$d_E(x, y) = \min \left\{ d_H(x', y') \mid (x', y') \text{ é alinhamento de } (x, y) \right\}.$$

O Problema da Distância de Edição (DE)

- Entrada: Um par (x, y) de palavras sobre um alfabeto (fixo) Σ.

Alinhamento ótimo: um alinhamento (x', y') de (x, y) é ótimo se $d_H(x', y')$ tem o menor valor possível.
(se existem vários possíveis, o ótimo deve minimizar $|x'|$)

Distância de Edição

A *Distância de Edição* (ou *Distância de Levenshtein*) entre duas palavras x e y é a distância de Hamming de um alinhamento ótimo de (x, y), isto é,

$$d_E(x, y) = \min \left\{ d_H(x', y') \mid (x', y') \text{ é alinhamento de } (x, y) \right\}.$$

O Problema da Distância de Edição (DE)

- Entrada: Um par (x, y) de palavras sobre um alfabeto (fixo) Σ.
- Saída: A distância de edição entre x e y.

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$.

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Observe que, para toda palavra x sobre Σ,

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Observe que, para toda palavra x sobre Σ,

$$x_{(0)} \quad = \quad \varepsilon,$$

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Observe que, para toda palavra x sobre Σ,

$$\begin{aligned} x_{(0)} &= \varepsilon, \\ x_{(|x|)} &= x, \end{aligned}$$

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Observe que, para toda palavra x sobre Σ,

$$\begin{aligned} x_{(0)} &= \varepsilon, \\ x_{(|x|)} &= x, \end{aligned}$$

Consequentemente, para toda palavra x sobre Σ,

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Observe que, para toda palavra x sobre Σ,

$$\begin{aligned} x_{(0)} &= \varepsilon, \\ x_{(|x|)} &= x, \end{aligned}$$

Consequentemente, para toda palavra x sobre Σ,

$$d_E(x, \varepsilon) = |x|,$$

Seja $x = (x_1, \ldots, x_n)$ sobre Σ e $i \in [0..n]$. Definimos o *i-ésimo prefixo de x* por

$$x_{(i)} = (x_1, \ldots, x_i).$$

Observe que, para toda palavra x sobre Σ,

$$\begin{aligned} x_{(0)} &= \varepsilon, \\ x_{(|x|)} &= x, \end{aligned}$$

Consequentemente, para toda palavra x sobre Σ,

$$\begin{aligned} d_E(x, \varepsilon) &= |x|, \\ d_E(x, y) &= d_E(x_{|x|}, y_{|y|}). \end{aligned}$$

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

- Caso 1: $\overline{x}_k \neq \star$ e $\overline{y}_k = \star$

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

- Caso 1: $\overline{x}_k \neq \star$ e $\overline{y}_k = \star$
  Neste caso $\overline{x}_k = x_i$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$,
  onde (x', y') é um alinhamento ótimo de $(x_{(i-1)}, y_{(j)})$.

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

- Caso 1: $\overline{x}_k \neq \star$ e $\overline{y}_k = \star$
  Neste caso $\overline{x}_k = x_i$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$,
  onde (x', y') é um alinhamento ótimo de $(x_{(i-1)}, y_{(j)})$.
  Consequentemente $d_E(x_{(i)}, y_{(j)}) \leq 1 + d_E(x_{(i-1)}, y_{(j)})$

Alinhamento de Sequências e Distância de Edição

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

- Caso 1: $\overline{x}_k \neq \star$ e $\overline{y}_k = \star$
  Neste caso $\overline{x}_k = x_i$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$,
  onde (x', y') é um alinhamento ótimo de $(x_{(i-1)}, y_{(j)})$.
  Consequentemente $d_E(x_{(i)}, y_{(j)}) \leq 1 + d_E(x_{(i-1)}, y_{(j)})$

- Caso 2: $\overline{x}_k = \star$ e $\overline{y}_k \neq \star$

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

- Caso 1: $\overline{x}_k \neq \star$ e $\overline{y}_k = \star$
  Neste caso $\overline{x}_k = x_i$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$,
  onde (x', y') é um alinhamento ótimo de $(x_{(i-1)}, y_{(j)})$.
  Consequentemente $d_E(x_{(i)}, y_{(j)}) \leq 1 + d_E(x_{(i-1)}, y_{(j)})$

- Caso 2: $\overline{x}_k = \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{y}_k = y_j$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$
  onde (x', y') é um alinhamento ótimo de $(x_{(i)}, y_{(j-1)})$.

Alinhamento de Sequências e Distância de Edição

Seja $(\overline{x}, \overline{y})$ um alinhamento ótimo de $(x_{(i)}, y_{(j)})$ e seja $k = |\overline{x}|$.

Há três possibilidades para $(\overline{x}_k, \overline{y}_k)$.

- Caso 1: $\overline{x}_k \neq \star$ e $\overline{y}_k = \star$
  Neste caso $\overline{x}_k = x_i$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$,
  onde (x', y') é um alinhamento ótimo de $(x_{(i-1)}, y_{(j)})$.
  Consequentemente $d_E(x_{(i)}, y_{(j)}) \leq 1 + d_E(x_{(i-1)}, y_{(j)})$

- Caso 2: $\overline{x}_k = \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{y}_k = y_j$ e $d_H(\overline{x}, \overline{y}) = 1 + d_H(x', y')$
  onde (x', y') é um alinhamento ótimo de $(x_{(i)}, y_{(j-1)})$.
  Consequentemente $d_E(x_{(i)}, y_{(j)}) \leq 1 + d_E(x_{(i)}, y_{(j-1)})$

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

  Seja (x', y') um alinhamento ótimo de $(x_{(i-1)}, y_{(j-1)})$ e os seguintes subcasos.

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

  Seja (x', y') um alinhamento ótimo de $(x_{(i-1)}, y_{(j-1)})$ e os seguintes subcasos.

  $\overline{x}_k \neq \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) \leq 1 + d_H(x', y')$

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

  Seja (x', y') um alinhamento ótimo de $(x_{(i-1)}, y_{(j-1)})$ e os seguintes subcasos.

  $\overline{x}_k \neq \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) \leq 1 + d_H(x', y')$
  $\overline{x}_k = \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) = d_H(x', y')$

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

  Seja (x', y') um alinhamento ótimo de $(x_{(i-1)}, y_{(j-1)})$ e os seguintes subcasos.

  $\overline{x}_k \neq \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) \leq 1 + d_H(x', y')$
  $\overline{x}_k = \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) = d_H(x', y')$

  Em resumo, temos

$$d_H(\overline{x}, \overline{y}) \leq d(\overline{x}_k, \overline{y}_k) + d_H(x', y') = d(x_i, y_j) + d_H(x', y').$$

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

  Seja (x', y') um alinhamento ótimo de $(x_{(i-1)}, y_{(j-1)})$ e os seguintes subcasos.

  $\overline{x}_k \neq \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) \leq 1 + d_H(x', y')$
  $\overline{x}_k = \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) = d_H(x', y')$

  Em resumo, temos

$$d_H(\overline{x}, \overline{y}) \leq d(\overline{x}_k, \overline{y}_k) + d_H(x', y') = d(x_i, y_j) + d_H(x', y').$$

  Consequentemente

- Caso 3: $\overline{x}_k \neq \star$ e $\overline{y}_k \neq \star$
  Neste caso $\overline{x}_k = x_i$ e $\overline{y}_k = y_j$

  Seja (x', y') um alinhamento ótimo de $(x_{(i-1)}, y_{(j-1)})$ e os seguintes subcasos.

  $\overline{x}_k \neq \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) \leq 1 + d_H(x', y')$
  $\overline{x}_k = \overline{y}_k$: neste caso temos $d_H(\overline{x}, \overline{y}) = d_H(x', y')$

  Em resumo, temos

  $$d_H(\overline{x}, \overline{y}) \leq d(\overline{x}_k, \overline{y}_k) + d_H(x', y') = d(x_i, y_j) + d_H(x', y').$$

  Consequentemente

  $$d_E(x_{(i)}, y_{(j)}) \leq d(x_i, y_j) + d_E(x_{(i-1)}, y_{(j-1)}).$$

Como só há estas três possibilidades, $d_E(x_{(i)}, y_{(j)})$ será o menor dentre

$$\begin{aligned}
&d_E(x_{(i-1)}, y_{(j)}) + 1,\\
&d_E(x_{(i)}, y_{(j-1)}) + 1, \text{ e}\\
&d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j).
\end{aligned}$$

Como só há estas três possibilidades, $d_E(x_{(i)}, y_{(j)})$ será o menor dentre

$$
\begin{aligned}
& d_E(x_{(i-1)}, y_{(j)}) + 1, \\
& d_E(x_{(i)}, y_{(j-1)}) + 1, \text{ e} \\
& d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j).
\end{aligned}
$$

Teorema

Se x e y são palavras sobre Σ e $i \in [1..|x|]$ e $j \in |y|$, então

$$
d_E(x_{(i)}, y_{(j)}) = \begin{cases} i + j, & \text{se } i = 0 \text{ ou } j = 0, \\ \min\left\{ d_E(x_{(i)}, y_{(j-1)}), d_E(x_{(i-1)}, y_{(j)}), d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j) \right\}, & \text{se } i > 0 \text{ e } j > 0. \end{cases}
$$

Teorema

Se x e y são palavras sobre Σ e $i \in [1..|x|]$ e $j \in |y|$, então

$$d_E(x_{(i)}, y_{(j)}) = \begin{cases} i + j, & \text{se } i = 0 \text{ ou } j = 0, \\ \min\left\{ d_E(x_{(i)}, y_{(j-1)}), d_E(x_{(i-1)}, y_{(j)}), d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j) \right\}, & \text{se } i > 0 \text{ e } j > 0. \end{cases}$$

Alinhamento de Sequências e Distância de Edição

Teorema

Se x e y são palavras sobre Σ e $i \in [1..|x|]$ e $j \in |y|$, então

$$d_E(x_{(i)}, y_{(j)}) = \begin{cases} i + j, & \text{se } i = 0 \text{ ou } j = 0, \\ \min\left\{ d_E(x_{(i)}, y_{(j-1)}), d_E(x_{(i-1)}, y_{(j)}), d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j) \right\}, & \text{se } i > 0 \text{ e } j > 0. \end{cases}$$

Em outras palavras:

- Nosso subproblemas serão prefixos de x e y

Teorema

Se x e y são palavras sobre Σ e $i \in [1..|x|]$ e $j \in |y|$, então

$$d_E(x_{(i)}, y_{(j)}) = \begin{cases} i + j, & \text{se } i = 0 \text{ ou } j = 0, \\ \min\left\{ d_E(x_{(i)}, y_{(j-1)}), d_E(x_{(i-1)}, y_{(j)}), d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j) \right\}, & \text{se } i > 0 \text{ e } j > 0. \end{cases}$$

Em outras palavras:

- Nosso subproblemas serão prefixos de x e y

$\text{DistEdit}_R(x, y)$

Se $|x| = 0$ *ou* $|y| = 0$
    Devolva $|x| + |y|$
$d_1 \leftarrow \text{DistEdit}_R(x, y_{(|y|-1)}) + 1$
$d_2 \leftarrow \text{DistEdit}_R(x_{(|x|-1)}, y) + 1$
Se $d_1 > d_2$
    $d_1 \leftarrow d_2$
$d_2 \leftarrow \text{DistEdit}_R(x_{(|x|-1)}, y_{(|y|-1)}) + d(x_{|x|}, y_{|y|})$
Se $d_1 > d_2$
    $d_1 \leftarrow d_2$
Devolva d_1

- Está correto:

Teorema

Se x e y são palavras sobre Σ e $i \in [1..|x|]$ e $j \in |y|$, então

$$d_E(x_{(i)}, y_{(j)}) = \begin{cases} i+j, & \text{se } i = 0 \text{ ou } j = 0, \\ \min\left\{d_E(x_{(i)}, y_{(j-1)}), d_E(x_{(i-1)}, y_{(j)}), d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j)\right\}, & \text{se } i > 0 \text{ e } j > 0. \end{cases}$$

Em outras palavras:

- Nosso subproblemas serão prefixos de x e y

$\text{DistEdit}_R(x, y)$

Se $|x| = 0$ *ou* $|y| = 0$
    Devolva $|x| + |y|$
$d_1 \leftarrow \text{DistEdit}_R(x, y_{(|y|-1)}) + 1$
$d_2 \leftarrow \text{DistEdit}_R(x_{(|x|-1)}, y) + 1$
Se $d_1 > d_2$
    $d_1 \leftarrow d_2$
$d_2 \leftarrow \text{DistEdit}_R(x_{(|x|-1)}, y_{(|y|-1)}) + d(x_{|x|}, y_{|y|})$
Se $d_1 > d_2$
    $d_1 \leftarrow d_2$
Devolva d_1

- Está correto: Recorrência

Teorema

Se x e y são palavras sobre Σ e $i \in [1..|x|]$ e $j \in |y|$, então

$$d_E(x_{(i)}, y_{(j)}) = \begin{cases} i + j, & \text{se } i = 0 \text{ ou } j = 0, \\ \min\left\{ d_E(x_{(i)}, y_{(j-1)}), d_E(x_{(i-1)}, y_{(j)}), d_E(x_{(i-1)}, y_{(j-1)}) + d(x_i, y_j) \right\}, & \text{se } i > 0 \text{ e } j > 0. \end{cases}$$

Em outras palavras:

- Nosso subproblemas serão prefixos de x e y

$\text{DistEdit}_R(x, y)$

Se $|x| = 0$ *ou* $|y| = 0$
  Devolva $|x| + |y|$
$d_1 \leftarrow \text{DistEdit}_R(x, y_{(|y|-1)}) + 1$
$d_2 \leftarrow \text{DistEdit}_R(x_{(|x|-1)}, y) + 1$
Se $d_1 > d_2$
  $d_1 \leftarrow d_2$
$d_2 \leftarrow \text{DistEdit}_R(x_{(|x|-1)}, y_{(|y|-1)}) + d(x_{|x|}, y_{|y|})$
Se $d_1 > d_2$
  $d_1 \leftarrow d_2$
Devolva d_1

- Está correto: Recorrência
- Quanto custa: $\Theta\left((1 + \sqrt{2})^n\right)$ onde $n = |x| + |y|$ (Exercício).

É possível fazer melhor?

É possível fazer melhor? Sim:

É possível fazer melhor? Sim:

$\text{DistEdit}_D(x, y)$

$D \leftarrow$ matriz indexada por$[0..|x|] \times [0..|y|]$
Para i *de* 0 *até* $|x|$
    $D[i, 0] \leftarrow i$
Para j *de* 0 *até* $|y|$
    $D[0, j] \leftarrow j$
Para i *de* 1 *até* $|x|$
    Para j *de* 1 *até* $|y|$
        Preenche(D, i, j)
Devolva $D[|x|, |y|]$

---

onde

Preenche(D, i, j)

$D[i, j] \leftarrow D[i-1, j] + 1$
$d' \leftarrow D[i, j-1] + 1$
Se $D[i, j] > d'$
    $D[i, j] \leftarrow d'$
$d' \leftarrow D[i, j] + d(x_i, y_j)$
Se $D[i, j] > d'$
    $D[i, j] \leftarrow d'$

- Está correto?

- Está correto? Tanto quanto a recorrência.

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\text{DistEdit}}(x, y) =$

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\mathrm{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\mathrm{DistEdit}}(x, y) = \Theta(1) + (|x| + 1)\Theta(1) + (|y| + 1)\Theta(1) + |x||y|\Theta(1) + \Theta(1) =$

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\text{DistEdit}}(x, y) = \Theta(1) + (|x| + 1)\Theta(1) + (|y| + 1)\Theta(1) + |x||y|\Theta(1) + \Theta(1) =$
$\Theta(1) + \Theta(|x|) + \Theta(|y|) + \Theta(|x||y|) =$

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\text{DistEdit}}(x, y) = \Theta(1) + (|x| + 1)\Theta(1) + (|y| + 1)\Theta(1) + |x||y|\Theta(1) + \Theta(1) = \Theta(1) + \Theta(|x|) + \Theta(|y|) + \Theta(|x||y|) = \Theta(|x||y|)$.

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\text{DistEdit}}(x, y) = \Theta(1) + (|x| + 1)\Theta(1) + (|y| + 1)\Theta(1) + |x||y|\Theta(1) + \Theta(1) = \Theta(1) + \Theta(|x|) + \Theta(|y|) + \Theta(|x||y|) = \Theta(|x||y|)$.

- É possível fazer melhor?

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\text{DistEdit}}(x, y) = \Theta(1) + (|x| + 1)\Theta(1) + (|y| + 1)\Theta(1) + |x||y|\Theta(1) + \Theta(1) = \Theta(1) + \Theta(|x|) + \Theta(|y|) + \Theta(|x||y|) = \Theta(|x||y|)$.

- É possível fazer melhor? Provavelmente não:

- Está correto? Tanto quanto a recorrência.
- Quanto custa?

Teorema

A execução de $\text{DistEdit}_D(x, y)$ toma tempo $\Theta(|x||y|)$.

Prova: Da análise direta do algoritmo temos
$T_{\text{DistEdit}}(x, y) = \Theta(1) + (|x| + 1)\Theta(1) + (|y| + 1)\Theta(1) + |x||y|\Theta(1) + \Theta(1) = \Theta(1) + \Theta(|x|) + \Theta(|y|) + \Theta(|x||y|) = \Theta(|x||y|)$.

- É possível fazer melhor? Provavelmente não:

Teorema (Backurs, Indyk 18)

Não existe algoritmo para DE com tempo $O(n^{2-\epsilon})$, onde $n = \max\{|x|, |y|\}$, para nenhum $\epsilon > 0$, a menos que a *Hipótese Forte do Tempo Exponencial* não seja verdadeira.