21.10 (1)

Def. Eine Editoperation ist ein Paar (α, β) mit $\alpha, \beta \in (A^1 \cup \{\varepsilon\})$ und $\alpha \neq \varepsilon$ oder $\beta \neq \varepsilon$

$\Rightarrow$ $\varepsilon \to \varepsilon$ ist keine Eop.

Schreibweise (α, β)

$$\alpha \longrightarrow \beta$$

Drei Möglichkeiten für $\alpha \to \beta$

- $\alpha \in A^1$, $\beta = \varepsilon$
  $\Rightarrow$ $a \longrightarrow \varepsilon$ für $a \in A$ Löschen von a
- $\alpha = \varepsilon$, $\beta \in A^1$
  $\Rightarrow$ $\varepsilon \to b$ für ein $b \in A$ Einfügen von b
- $\alpha, \beta \in A^1$
  $\Rightarrow$ $a \longrightarrow b$ für $a, b \in A$ Ersetzung von a durch b

Löschen oder Einfügen = indel.

Def Ein Alignment A von Sequenzen u und v ist eine Folge $(\alpha_1 \to \beta_1, \alpha_2 \to \beta_2, \ldots, \alpha_h \to \beta_h)$ mit

$$\alpha_1 \alpha_2 \ldots \alpha_h = u$$
$$\beta_1 \beta_2 \ldots \beta_h = v$$

Beobachtung: 21.10. (2)

Seien u und v Sequenzen der Längen m bzw. n.
Sei A ein Alignment von u und v
Dann gilt $m+n \geq |A| = h \geq \max\{m,n\}$

Die Anzahl der Alignments von zwei festen aber beliebigen Sequenzen der Längen m und n.

- 1. $m=0, n=0$
  - Sei A ein Alignment von u und v der Länge h

$m+n \geq h \geq \max\{m,n\}$

$0+0 \geq h \geq \max\{0,0\} = 0$

$\Rightarrow h=0$

A ist das leere Alignment, ~~[illegible]~~

$\underline{1}$

2. $m=0$ und $n>0$

$m+n = 0+n = n \geq h \geq \max\{m,n\} = n$

A hat die Länge $h=n$

$\underline{1}$

3. $m>0$ und $n=0$ $\underline{1}$

$m>0,\ n>0$ 21.10. 3

Jedes Alignment von u und v enthält ~~stets~~ mindestens eine Editoperation. Betrachte die letzte Editop. in den Alignments von u und v und zerlege die Menge aller Alignments in drei Teilmengen

- betrachte die Menge der Alignments von u und v, die mit Löschung enden
  $\Rightarrow$ alle ~~Entt.~~ enthalten die Löschop. $u[m] \to \varepsilon$
  $\Rightarrow$ diese Menge ist genauso groß wie die Menge der Alignments die $u[1..m-1]$ und v alignieren

  $Aligns(m-1, n)$

- betrachte Menge der Alignments von u und v, die mit Einfügung enden

  Einfügung $\varepsilon \to v[n]$

  $\Rightarrow$ Menge ist genauso groß wie Menge der Alignments von u und $v[1..n-1]$
  $Aligns(m, n-1)$

21.10 4

$\text{Aligns}(m,n) = 1$ für $m=0$ oder $n=0$

$\text{Aligns}(m,n) =$ für $m>0$ und $n>0$

$\text{Aligns}(m-1,n) +$
$\text{Aligns}(m,n-1) +$
$\text{Aligns}(m-1,n-1)$

---

3. Fall

$\Rightarrow \text{Aligns}(m-1, n-1)$

$$\text{Aligns}(n,n) \approx (1+\sqrt{2})^{2n+1} \cdot \sqrt{n}$$

$n = 1000 \Rightarrow (1+\sqrt{2})^{2001} \cdot \sqrt{1000} = 10^{767,4}$

$$\begin{pmatrix} a & - \\ - & b \end{pmatrix} \qquad \begin{pmatrix} - & a \\ b & - \end{pmatrix}$$

<u>Def</u> Eine Subsequenz von u und v ist eine Folge von Index Paaren

$(i_1, j_1), (i_2, j_2), \ldots, (i_r, j_r)$

mit

$1 \leq i_1 < i_2 \ldots < i_r \leq m$

$1 \leq j_1 < j_2 \ldots < j_r \leq n$

(i_k, j_k)

$u[i_k]$

$\downarrow$

$v[j_k]$

21.10. 5

Beobachtung Sei Subseqs(m,n) die Anzahl der Subsequenzen von zwei Sequenzen der Längen m und n. Es gilt:

$$\mathrm{Subseqs}(m,n) = \sum_{r=0}^{\min\{m,n\}} \binom{m}{r} \cdot \binom{n}{r}$$

Approx

$$\mathrm{Subseqs}(n,n) = 2^{2n} \cdot (4 \cdot \sqrt{n\pi})^{-1}$$

$n=1000$

$$\mathrm{Subseqs}(1000,1000) = 10^{600}$$

$$\binom{m}{r} = \frac{m!}{r!\,(m-r)!} \text{ für } m \geq r$$

Bewertungsmodelle für Alignments

Def: Eine Kostenfunktion δ weist einer Editop. $\alpha \rightarrow \beta$, $\alpha \neq \beta$ positive reelle Kosten $\underbrace{\delta(\alpha \rightarrow \beta)}_{\text{Kosten von } \alpha \rightarrow \beta}$ zu

Die Kosten von $\alpha \rightarrow \alpha$ sind $\delta(\alpha \rightarrow \alpha) = 0$

Einheitskosten:

$$\delta(\alpha \rightarrow \beta) = \begin{cases} 0 & , \alpha = \beta \\ 1 & , \text{sonst} \end{cases}$$

21.10. ⑥

Sei $A = (\alpha_1 \to \beta_1, \dots, \alpha_h \to \beta_h)$ ein Alignment. Sei δ eine Kostenfunktion. Die Kosten $\delta(A)$ von A sind def. durch

$$\delta(A) = \sum_{i=1}^{h} \delta(\alpha_i \to \beta_i)$$

<u>Def.</u> Die Edit Distanz von u und v, bezeichnet durch $edist_\delta(u,v)$ ist

$$edist_\delta(u,v) = \min\{\delta(A) \mid A \text{ ist ein Alignment von } u \text{ und } v\} \quad (*)$$

A ist ein optimales Alignment von u und v falls $\delta(A) = edist_\delta(u,v)$.

Problemdef.:

ED-Problem:

- berechne $edist_\delta(u,v)$
- bestimme die opt. Alignments