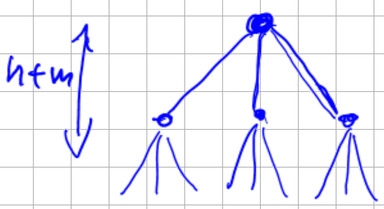
Minimale Editierdistanz

DNA-Vergleich
Textabstände (Auto-Korrektur)

ZIEGE (n)
↓
TIGER (m)

A Einfügen ①
B Entfernen ①
C Ersetzen ① / Beibehalten ⓪

ZIEGE (n)	A	ZIEGE|R TIGE|R	⟹	ZIEGE (n) TIGE (m−1) ①
TIGER (m)	B	ZIEG TIGER	⟹	ZIEG (n−1) TIGER (m) ①
	C	ZIEG|R TIGE|R	⟹	ZIEG (n−1) TIGE (m−1) ①



$n=0 \Rightarrow$ Kosten m

$m=0 \Rightarrow$ Kosten n

gegeben: Zeichenketten $A_n=(a_1,\dots,a_n)$
$B_m=(b_1,\dots,b_m)$ $\quad n,m\in\mathbb{N}$

gesucht: günstigste zeichenweise Transformation

$$(a_1,\dots,a_n)\to(b_1,\dots,b_m)$$

mit Kosten

Operation	Levenshtein	Längste gemeinsame Teilfolge	allg.
c einfügen	1	1	ins(c)
c löschen	1	1	del(c)
ersetzen $c\to c'$	$\mathbb{1}\{c\neq c'\}$	$\infty\cdot\mathbb{1}\{c\neq c'\}$	repl(c, c')

⓪ $E(n,m)$: minimale Editierdistanz

für $(a_1, \dots, a_n)$, $(b_1, \dots, b_m)$

① Teilprobleme?

$E(i,j)$, ED von $(a_1, \dots, a_i)$, $(b_1, \dots, b_j)$

$0 \le i \le n$, $0 \le j \le m$ $\quad$ $\# TP = n \cdot m$

② Raten / Probieren:

(a) Löschen $\quad a_{1,\dots,i} \to a_{1,\dots,i-1}$

(b) Einfügen $\quad b_{1,\dots,j} \to b_{1,\dots,j-1}$

(c) Ersetzen $\quad a_{1,\dots,i} \to a_{1 \dots i-1}$, $\quad b_{1,\dots,j} \to b_{1 \dots j-1}$

③ Rekursion:

$$E(i,j) = \min \begin{cases} del(a_i) + E(i-1,j) \\ ins(b_j) + E(i,j-1) \\ repl(a_i,b_j) + E(i-1,j-1) \end{cases}$$

④ Abhängigkeiten

$$\begin{tikzcd} E(i-1,j-1) & E(i,j-1) \\ E(i-1,j) & E(i,j) \arrow[l] \arrow[u] \arrow[ul] \end{tikzcd}$$

$\frac{\text{Kosten}}{\text{Teilproblem}} = \Theta(1)$

⑤ Lösung in $E(n,m)$ $\quad \Rightarrow$ Gesamtkosten $n \cdot m$

Beispiel FISCH $\longrightarrow$ FROSCH

	∅	F	I	S	C	H
∅	0	1	2	3	4	5
F	1	0	1	2	3	4
R	2	1	1	2	3	4
O	3	2	2	2	3	4
S	4	3	3	2	3	4
C	5	4	4	3	2	3
H	6	5	5	4	3	2

$\leftarrow$ löschen

$\uparrow$ einfügen

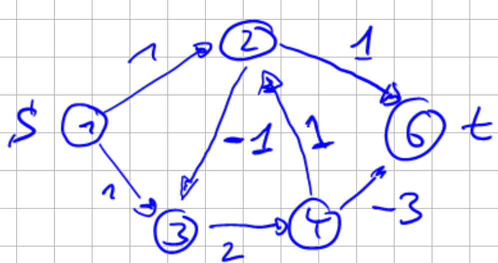
$\nwarrow$ ersetzen

1. $I \to R$, $+O$
2. $+R$, $O \to I$

Bellman-Ford Algorithmus

gegeben Graph $G = (V, E, c)$, $\quad c \in \mathbb{R}^E$

gesucht kürzester Pfad von $s \in V$ nach $t \in V$



⓪ gesucht:

Kürzeste Weglänge $d(s,t)$

① Teilprobleme

$d_\ell(s,t)$: Kürzeste Weglänge von $s \to t$ mit ℓ Kanten

② Aufzählen

Schlusskante nach v bei ℓ Kanten

③ Rekursion

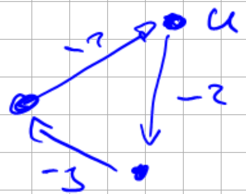
$$d_{\underline{\underline{\ell}}}(s,v) = \min \left\{ d_{\ell-1}(s,u) + c(u,v),\ u \in N^-(v) \right\}$$

$$d_0(s,s) = 0, \quad d_0(s,v) = \infty \ \forall v \neq s$$

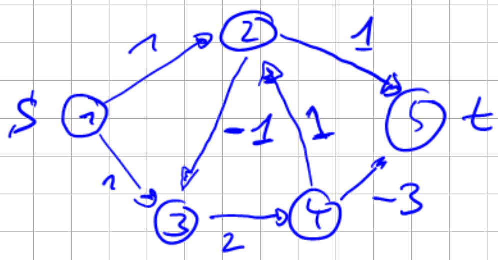
④ Abhängigkeiten

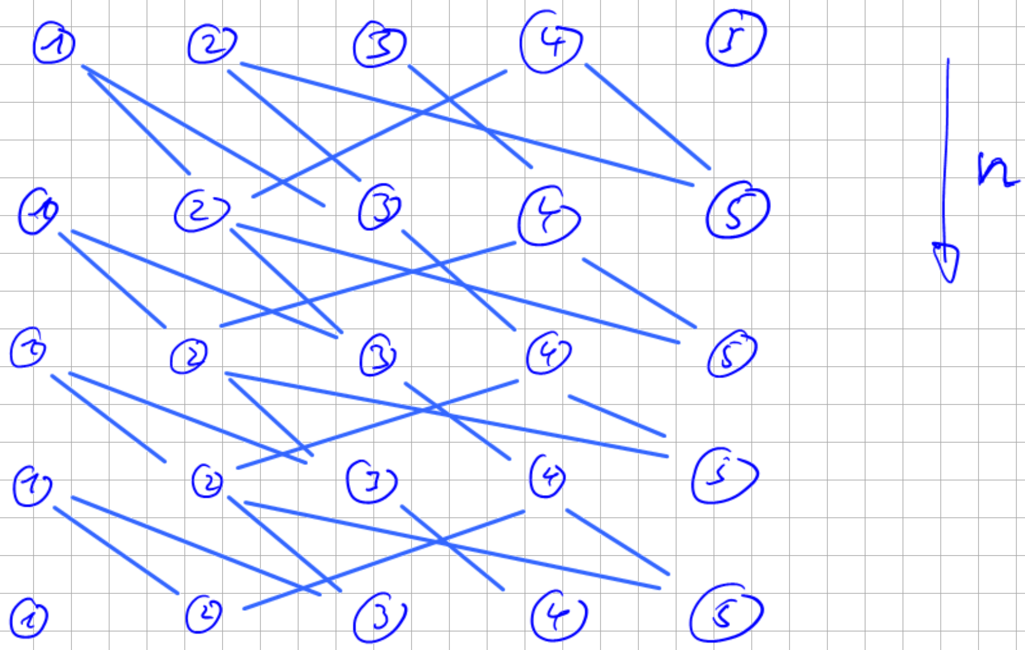
$d_\ell(s,u)$ hängt von allen Vorgängern von u ab

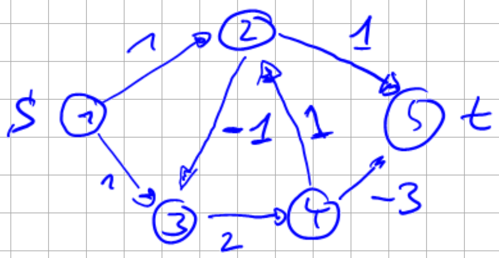
Zyklen?



⑤ Lösung: $d_n(s,t)$ [oder Zyklus]







	1	2	3	4	5
1		1	1		
2			-1		1
3				2	
4		1			-3
5					

l	1	2	3	4	5 → n	
0	0	∞	∞	∞	∞	
1	0	1	1	∞	∞	
2	0	1	0	3	2	
3	0	1	0	2	0	
4	0	1	0	2	-1	
5	0	1	0	2	-1	← keine Änderung