Poravnanje nizova

Seminar iz kolegija *Oblikovanje i analiza algoritama*

Goran Flegar

Prirodoslovno-matematički fakultet
Sveučilište u Zagrebu

20. siječnja 2015.

Sadržaj

Definicija problema

Algoritam za optimalno poravnanje

Algoritam prostorne složenosti $O(n+m)$

Sadržaj

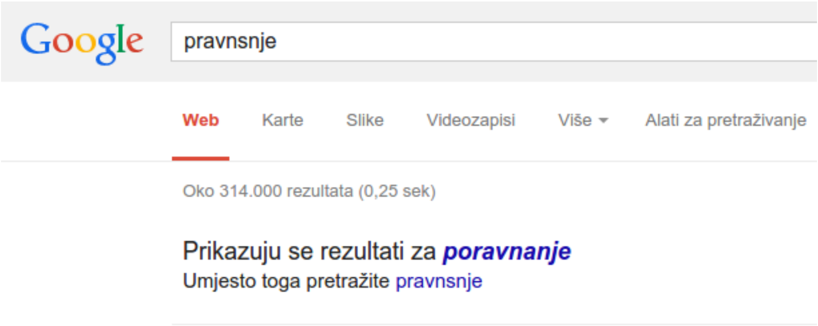
Definicija problema

Algoritam za optimalno poravnanje

Algoritam prostorne složenosti $O(n + m)$

Motivacija

- Ispravljanje pogrešnog unosa korisnika
  - Potrebno je u bazi riječi pronaći onu "najsličniju" zadanoj



Motivacija (2)

- Primjene u bioinformatici
  - Određivanje srodstva između dvije vrste na temelju sličnosti njihovih DNA
  - *Needleman-Wunsch* algoritam

```
GAATTCAG     GAATTCAG
| | || |     | || | |
GGA-TC-G     GCAT-C-G

GAATTC-A     GAATTC-A
| | || |     | || | |
GGA-TCGA     GCAT-CGA
```

Definicija poravnanja

- Zašto su "*pravsnje*" i "*poravnanje*" slični?
  - Možemo ih "poravnati" na sljedeći način uz mali broj "razlika"

$$P = \begin{array}{c|c|cccc|c|ccc} \mathrm{p} & \text{-} & \mathrm{r} & \mathrm{a} & \mathrm{v} & \mathrm{n} & \mathrm{s} & \mathrm{n} & \mathrm{j} & \mathrm{e} \\ \mathrm{p} & \mathrm{o} & \mathrm{r} & \mathrm{a} & \mathrm{v} & \mathrm{n} & \mathrm{a} & \mathrm{n} & \mathrm{j} & \mathrm{e} \end{array}$$

Definicija

Neka su zadani nizovi $x = x_1 x_2 \dots x_n$ *i* $y = y_1 y_2 \dots y_m$, $(x_i, y_j \in A$, *A je zadani alfabet). Definiramo skupove indeksa* $I := \{1 \dots n\}$, $J := \{1 \dots m\}$. *Za skup* $P \subseteq I \times J$ *kažemo da je poravnanje nizova* x *i* y *ako vrijedi sljedeće:*

1. $\forall (i,j), (i',j') \in P \quad i = i' \Leftrightarrow j = j'$
   - *svaki element jednog niza poravnat je s najviše jednim elementom drugog niza*
2. $\forall (i,j), (i',j') \in P \quad i < i' \Leftrightarrow j < j'$
   - *nema "unakrsnog" poravnavanja, tj. zamjene poretka elemenata u nizovima*

- $P = \{(1,1), (2,3), (3,4), (4,5), (5,6), (6,7), \dots, (9,10)\}$

Definicija poravnanja (2)

- Promotrimo sljedeća dva poravnanja

$$P_1 = \begin{array}{c|c|cccc|c|ccc} p & - & r & a & v & n & s & n & j & e \\ p & o & r & a & v & n & a & n & j & e \end{array}$$

$$P_2 = \begin{array}{c|c|c|c|c|c|c|ccc} p & r & a & v & n & s & - & n & j & e \\ p & o & r & a & v & n & a & n & j & e \end{array}$$

- 1. poravnanje izgleda bolje
  - Treba nam neka mjera "dobrote" poravnanja

Definicija

Neka je zadana cijena nepodudaranja $\alpha : A \times A \to \mathbb{R}$ *(uobičajeno je* $\alpha_{pp} = 0$*) i kazna praznine* $\delta > 0$. *Definiramo cijenu poravnanja* P *s*

$$c(P) = \sum_{(i,j) \in P} \alpha_{x_i y_j} + \delta(n + m - 2|P|).$$

- Ako su zadani $\alpha_{pq} = 1 - \delta_{pq}$ (δ_{ij} je Kroneckerov simbol) i $\delta = 1$ tada je $c(P_1) = 2$ i $c(P_2) = 6$

Problem poravnanja nizova

- Za zadane nizove x i y, cijenu nepodudaranja α i kaznu praznine δ potrebno je u skupu svih poravnanja $\mathcal{P}$ nizova x i y pronaći optimalno poravnanje P, tj. poravnanje za koje vrijedi $c(P) = \min\{c(P')|P' \in \mathcal{P}\}$.

Sadržaj

Definicija problema

Algoritam za optimalno poravnanje

Algoritam prostorne složenosti $O(n+m)$

Svojstva poravnanja

Lema

Neka je $P \subseteq I \times J$ poravnanje nizova $x = x_1 x_2 \ldots x_n$ i $y = y_1 y_2 \ldots y_m$. Tada vrijedi barem jedno od sljedećeg:

1. $(n, m) \in P$
2. $(\{n\} \times J) \cap P = \emptyset$
3. $(I \times \{m\}) \cap P = \emptyset$.

Dokaz.

Jedina preostala mogućnost je $(n, j), (i, m) \in P$ za neke $j < m$ i $i < n$, što je kontradikcija s drugim svojstvom iz definicije poravnanja. $\square$

Svojstva poravnanja (2)

Teorem

Neka je zadan problem poravnanja nizova $(x_1 \ldots x_n, y_1 \ldots y_m, \alpha, \delta)$,
Neka P_{ij} *rješenje problema* $(x_1 \ldots x_i, y_1 \ldots y_j, \alpha, \delta)$ *i*
$o(i,j) := c(P_{ij})$, $i = 0 \ldots n$, $j = 0 \ldots m$. *Tada je*

$$o(i,j) = \min\{o(i-1,j) + \delta, o(i,j-1) + \delta, o(i-1,j-1) + \alpha_{x_i y_j}\}$$
$$o(i,0) = i\delta, \quad o(0,j) = j\delta$$

za sve $i = 1 \ldots n$, $j = 1 \ldots m$.
Nadalje, vrijedi sljedeće:

1. *Ako je* $(i,j) \in P_{ij}$ *onda je* $P_{ij} \setminus \{(i,j)\}$ *rješenje za* $(x_1 \ldots x_{i-1}, y_1 \ldots y_{j-1}, \alpha, \delta)$.
2. *Ako je* $(\{i\} \times J) \cap P_{ij} = \emptyset$ *onda je* P_{ij} *rješenje za* $(x_1 \ldots x_{i-1}, y_1 \ldots y_j, \alpha, \delta)$.
3. *Ako je* $(I \times \{j\}) \cap P_{ij} = \emptyset$ *onda je* P_{ij} *rješenje za* $(x_1 \ldots x_i, y_1 \ldots y_{j-1}, \alpha, \delta)$.

Svojstva poravnanja (3)

Ideja dokaza.

1. Kada bi postojalo poravnanje $Q_{i-1,j-1}$ nizova $x_1 \ldots x_{i-1}$ i $y_1 \ldots y_{j-1}$ takvo da je $c(Q_{i-1,j-1}) < c(P_{ij} \setminus \{(i,j)\})$, vrijedilo bi

   $$c(Q_{i-1,j-1} \cup \{(i,j)\}) < c(P_{ij})$$

   što je kontradikcija s činjenicom da je P_{ij} optimalno poravnanje. Sada iz $c(P_{ij}) = c(P_{ij} \setminus \{(i,j)\}) + \alpha_{x_i y_j}$ slijedi $o(i,j) = o(i-1, j-1) + \alpha_{x_i y_j}$.

Analogno se pokažu tvrdnje 2. i 3. te da u tom slučaju vrijedi formula $o(i,j) = o(i-1,j) + \delta$, odnosno $o(i,j) = o(i,j-1) + \delta$. Kako je svaki od tih izraza cijena nekog poravnanja, a optimalno poravnanje zadovoljava barem jedan od tih izraza, vrijedi rekurzija iz iskaza teorema. $\square$

Algoritam

- Na temelju rekurzije iz teorema možemo sastaviti algoritam za računanje $o(i,j)$.
  - Napravimo $(n+1) \times (m+1)$ tablicu u koju ćemo na ij-to mjesto zapisati vrijednost $o(i,j)$
  - Popunimo prvi redak i prvi stupac tablie koristeći početne uvjete rekurzije
  - Popunimo ostatak tablice redak po redak (ili stupac po stupac) koristeći rekurziju
    - Kada računamo $o(i,j)$ u tablici su već izračunati $o(i,j-1)$, $o(i-1,j)$ i $o(i-1,j-1)$.
- Složenost?
  - Vremenska: $\Theta(nm)$
  - Prostorna: $\Theta(nm)$

Primjer

Pronađite optimalno poravnanje nizova "nizovi" i "izkvui" uz parametre $\alpha_{pq} = 1 - \delta_{pq}$ i $\delta = 1$.

Primjer

$$o(i,0) = i\delta = i$$
$$o(0,j) = j\delta = j$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	?	?	?	?	?	?
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(1,1) = \min\{o(1,0)+\delta, o(0,1)+\delta, o(0,0)+\alpha_{ni}\}$$
$$= \min\{1+1, 1+1, 0+1\} = 1$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	?	?	?	?	?
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$\begin{aligned} o(1,2) &= \min\{o(1,1)+\delta, o(0,2)+\delta, o(0,1)+\alpha_{nz}\} \\ &= \min\{1+1, 2+1, 1+1\} = 2 \end{aligned}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	?	?	?	?
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(1,3) &= \min\{o(1,2)+\delta, o(0,3)+\delta, o(0,2)+\alpha_{nk}\} \\
&= \min\{2+1, 3+1, 2+1\} = 3
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	?	?	?
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(1,4) = \min\{o(1,3)+\delta, o(0,4)+\delta, o(0,3)+\alpha_{nv}\}$$
$$= \min\{3+1, 4+1, 3+1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	?	?
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(1,5) = \min\{o(1,4)+\delta, o(0,5)+\delta, o(0,4)+\alpha_{nu}\}$$
$$= \min\{4+1, 5+1, 4+1\} = 5$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	?
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(1,6) = \min\{o(1,5) + \delta, o(0,6) + \delta, o(0,5) + \alpha_{ni}\}$$
$$= \min\{5 + 1, 6 + 1, 5 + 1\} = 6$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	?	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(2,1) = \min\{o(2,0)+\delta, o(1,1)+\delta, o(1,0)+\alpha_{ii}\}$$
$$= \min\{2+1, 1+1, 1+0\} = 1$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	?	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(2,2) = \min\{o(2,1)+\delta, o(1,2)+\delta, o(1,1)+\alpha_{iz}\}$$
$$= \min\{1+1, 2+1, 1+1\} = 2$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	?	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(2,3) &= \min\{o(2,2)+\delta, o(1,3)+\delta, o(1,2)+\alpha_{ik}\} \\
&= \min\{2+1, 3+1, 2+1\} = 3
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	?	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(2,4) = \min\{o(2,3)+\delta, o(1,4)+\delta, o(1,3)+\alpha_{iv}\}$$
$$= \min\{3+1, 4+1, 3+1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	?	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(2,5) = \min\{o(2,4) + \delta, o(1,5) + \delta, o(1,4) + \alpha_{iu}\}$$
$$= \min\{4 + 1, 5 + 1, 4 + 1\} = 5$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	?
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(2,6) &= \min\{o(2,5)+\delta, o(1,6)+\delta, o(1,5)+\alpha_{ii}\} \\
&= \min\{5+1, 6+1, 5+0\} = 5
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	?	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(3,1) = \min\{o(3,0)+\delta, o(2,1)+\delta, o(2,0)+\alpha_{zi}\}$$
$$= \min\{3+1, 1+1, 2+1\} = 2$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	?	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$\begin{aligned} o(3,2) &= \min\{o(3,1)+\delta, o(2,2)+\delta, o(2,1)+\alpha_{zz}\} \\ &= \min\{2+1, 2+1, 1+0\} = 1 \end{aligned}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	?	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(3,3) &= \min\{o(3,2)+\delta, o(2,3)+\delta, o(2,2)+\alpha_{zk}\} \\
&= \min\{1+1, 3+1, 2+1\} = 2
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	?	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(3,4) = \min\{o(3,3)+\delta, o(2,4)+\delta, o(2,3)+\alpha_{zv}\}$$
$$= \min\{2+1, 4+1, 3+1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	?	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(3,5) = \min\{o(3,4)+\delta, o(2,5)+\delta, o(2,4)+\alpha_{zu}\}$$
$$= \min\{3+1, 5+1, 4+1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	?
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(3,6) &= \min\{o(3,5)+\delta, o(2,6)+\delta, o(2,5)+\alpha_{zi}\} \\
&= \min\{4+1, 5+1, 5+1\} = 5
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	?	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(4,1) = \min\{o(4,0)+\delta, o(3,1)+\delta, o(3,0)+\alpha_{oi}\}$$
$$= \min\{4+1, 2+1, 3+1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	?	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(4,2) = \min\{o(4,1)+\delta, o(3,2)+\delta, o(3,1)+\alpha_{oz}\}$$
$$= \min\{3+1, 1+1, 2+1\} = 2$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	?	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(4,3) &= \min\{o(4,2)+\delta, o(3,3)+\delta, o(3,2)+\alpha_{ok}\} \\
&= \min\{2+1, 2+1, 1+1\} = 2
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	?	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(4,4) &= \min\{o(4,3)+\delta, o(3,4)+\delta, o(3,3)+\alpha_{ov}\} \\
&= \min\{2+1, 3+1, 2+1\} = 3
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	?	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(4,5) = \min\{o(4,4) + \delta, o(3,5) + \delta, o(3,4) + \alpha_{ou}\}$$
$$= \min\{3 + 1, 4 + 1, 3 + 1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	?
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(4,6) = \min\{o(4,5)+\delta, o(3,6)+\delta, o(3,5)+\alpha_{oi}\}$$
$$= \min\{4+1, 5+1, 4+1\} = 5$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	?	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(5,1) = \min\{o(5,0)+\delta, o(4,1)+\delta, o(4,0)+\alpha_{vi}\}$$
$$= \min\{5+1, 3+1, 4+1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	?	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(5,2) = \min\{o(5,1) + \delta, o(4,2) + \delta, o(4,1) + \alpha_{vz}\}$$
$$= \min\{4 + 1, 2 + 1, 3 + 1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	?	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(5,3) = \min\{o(5,2) + \delta, o(4,3) + \delta, o(4,2) + \alpha_{vk}\}$$
$$= \min\{3 + 1, 2 + 1, 2 + 1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	?	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(5,4) = \min\{o(5,3)+\delta, o(4,4)+\delta, o(4,3)+\alpha_{vv}\}$$
$$= \min\{3+1, 3+1, 2+0\} = 2$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	?	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(5,5) = \min\{o(5,4)+\delta, o(4,5)+\delta, o(4,4)+\alpha_{vu}\}$$
$$= \min\{2+1, 4+1, 3+1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	?
6	i	6	?	?	?	?	?	?

Primjer

$$o(5,6) = \min\{o(5,5) + \delta, o(4,6) + \delta, o(4,5) + \alpha_{vi}\}$$
$$= \min\{3+1, 5+1, 4+1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	?	?	?	?	?	?

Primjer

$$o(6,1) = \min\{o(6,0)+\delta, o(5,1)+\delta, o(5,0)+\alpha_{ii}\}$$
$$= \min\{6+1, 4+1, 5+0\} = 5$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	?	?	?	?	?

Primjer

$$
\begin{aligned}
o(6,2) &= \min\{o(6,1)+\delta, o(5,2)+\delta, o(5,1)+\alpha_{iz}\} \\
&= \min\{5+1, 3+1, 4+1\} = 4
\end{aligned}
$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	?	?	?	?

Primjer

$$o(6,3) = \min\{o(6,2)+\delta, o(5,3)+\delta, o(5,2)+\alpha_{ik}\}$$
$$= \min\{4+1, 3+1, 3+1\} = 4$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	?	?	?

Primjer

$$o(6,4) = \min\{o(6,3)+\delta, o(5,4)+\delta, o(5,3)+\alpha_{iv}\}$$
$$= \min\{4+1, 2+1, 3+1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	?	?

Primjer

$$o(6,5) = \min\{o(6,4) + \delta, o(5,5) + \delta, o(5,4) + \alpha_{iu}\}$$
$$= \min\{3 + 1, 3 + 1, 2 + 1\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	?

Primjer

$$o(6,6) = \min\{o(6,5) + \delta, o(5,6) + \delta, o(5,5) + \alpha_{ii}\}$$
$$= \min\{3 + 1, 4 + 1, 3 + 0\} = 3$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{\}$$

$$3 = \min\{\ 4\ ,\ 5\ ,3\}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{(6,6)\}$$

$$3 = \min\{3, \; 5 \;, \; 4 \;\}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{(6,6)\}$$

$$2 = \min\{\ 4\ ,\ 4\ ,\ 2\ \}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{(5,4),(6,6)\}$$

$$2 = \min\{\ 3\ ,\ 3\ ,\ 2\}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{(4,3),(5,4),(6,6)\}$$

$$1 = \min\{\ 3\ ,\ 3\ ,1\}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{(3,2),(4,3),(5,4),(6,6)\}$$

$$1 = \min\{\ 3\ ,\ 2\ ,\ 1\}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

$$P = \{(2,1),(3,2),(4,3),(5,4),(6,6)\}$$

		0	1	2	3	4	5	6
			i	z	k	v	u	i
0		0	1	2	3	4	5	6
1	n	1	1	2	3	4	5	6
2	i	2	1	2	3	4	5	5
3	z	3	2	1	2	3	4	5
4	o	4	3	2	2	3	4	5
5	v	5	4	3	3	2	3	4
6	i	6	5	4	4	3	3	3

Primjer

Primjer (nastavak)

Optimalno poravnanje je $P = \{(2,1),(3,2),(4,3),(5,4),(6,6)\}$ *i* $o(6,6) = c(P) = 3$.

$$\begin{array}{c|cc|c|c|c|c}
n & i & z & o & v & - & i \\
- & i & z & k & v & u & i
\end{array}$$

Može i s manje memorije!

- Cijenu optimalnog poravnanja možemo naći s $\Theta(m)$ memorije
  - Računanje i-tog retka matrice ovisi samo o $(i-1)$-vom retku
  - Trebamo spremati samo dva retka matrice – trenutni i prethodni
- Problem
  - Nemamo dovoljno informacija da rekonstruiramo samo poravnanje

Sadržaj

Definicija problema

Algoritam za optimalno poravnanje

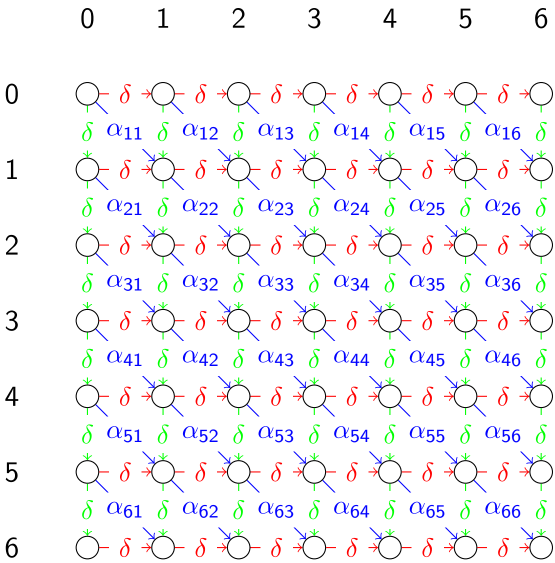
Algoritam prostorne složenosti $O(n+m)$

Još malo teorije

- Problemu poravnanja nizova (x, y, α, δ) pridružujemo usmjeren utežen graf $G = (V, E, \omega)$
  - $V = (I \cup \{0\}) \times (J \cup \{0\})$
  - $E = E_1 \cup E_2 \cup E_3$
    - $E_1 = \big\{((i-1, j-1), (i, j)) \;\big|\; i \in I, j \in J\big\}$
    - $E_2 = \big\{((i, j-1), (i, j)) \;\big|\; i \in I, j \in J\big\}$
    - $E_3 = \big\{((i-1, j), (i, j)) \;\big|\; i \in I, j \in J\big\}$
  - $\omega : E \to \mathbb{R}$

$$\omega(x) = \begin{cases} \delta & , x \in E_2 \cup E_3 \\ \alpha_{x_i y_j} & , x = ((i-1, j-1), (i, j)) \end{cases}$$

Pridruženi graf



Veza poravnanja i pridruženog grafa

Teorem

Označimo s $f(i,j)$ *najkraći put od čvora* $(0,0)$ *do* (i,j)*. Tada vrijedi*

$$f(i,j) = o(i,j), \quad i = 0,\ldots,n, j = 0,\ldots,m.$$

Posebno, cijena optimalnog poravnanja jednaka je duljini najkraćeg puta od $(0,0)$ *do* (n,m)*.*

Ideja dokaza.

Zadnji brid na najkraćem putu do (i,j) je brid iz jednog od vrhova $(i-1,j),(i,j-1),(i-1,j-1)$, pa je duljina najkraćeg puta $f(i,j) = \min\{f(i-1,j)+\delta, f(i,j-1)+\delta, f(i-1,j-1)+\alpha_{x_i y_j}\}$, $(i,j) \in I \times J$. Očito je $f(i,0) = i\delta, \quad i = 0,\ldots,n$ i $f(0,j) = j\delta, \quad j = 0,\ldots,m$, pa f zadovoljava istu rekurziju i iste početne uvjete kao i o. Indukcijom po $i+j$ se pokaže da vrijedi $f(i,j) = o(i,j)$. $\square$

Možemo okrenuti problem

- Označimo s $g(i,j)$ najkraći put od (i,j) do (n,m)
  - Prvi brid na tom putu je brid do jednog od vrhova $(i+1,j),(i,j+1),(i+1,j+1)$
  - Vrijedi rekurzija: $g(i,j) = \min\{g(i+1,j)+\delta, g(i,j+1)+\delta, g(i+1,j+1)+\alpha_{x_{i+1}y_{j+1}}\}$
  - Početni uvjeti: $g(i,m) = (n-i)\delta, \quad g(n,j) = (m-j)\delta$
- Možemo sastaviti algoritam za računanje $g(i,j)$ analogan onom za računanje $o(i,j)$
  - Verziju koja koristi $\Theta(m)$ memorije

Svojstva f i g

Teorem

Najkraći put od $(0,0)$ *do* (n,m) *koji prolazi kroz* (i,j) *ima duljinu* $l_{ij} = f(i,j) + g(i,j)$.

Dokaz.

Svaki put od $(0,0)$ do (n,m) koji prolazi kroz (i,j) se sastoji od nekog puta iz $(0,0)$ do (i,j) i nekog puta od (i,j) do (n,m) pa je $l_{ij} \geq f(i,j) + g(i,j)$. S druge strane, put koji se sastoji od najkraćeg puta iz $(0,0)$ do (i,j) i najkraćeg puta od (i,j) do (n,m) je jedan takav put pa je $l_{ij} \leq f(i,j) + g(i,j)$. $\square$

Svojstva f i g (2)

Teorem

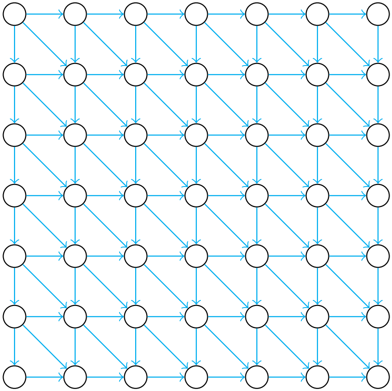
Za svaki $i \in I \cup \{0\}$ *postoji* $q \in J \cup \{0\}$ *takav da najkraći put od* $(0,0)$ *do* (n,m) *prolazi kroz* (i,q).
Posebno, vrijedi $f(n,m) = \min\{f(i,q) + g(i,q) \mid q \in J \cup \{0\}\}$.

Ideja dokaza.

Iz definicije pridruženog grafa vidimo da svaki put od $(0,0)$ do (n,m) prolazi kroz barem jedan čvor u svakom retku grafa (jedini način da dođemo u i-ti redak je iz $(i-1)$-vog retka, formalnije – indukcijom), pa tada i najkraći put prolazi kroz neki čvor i-tog retka grafa. Dakle, q iz teorema postoji. Najkraći put u cijelom grafu je ujedno i najkraći put koji prolazi kroz (i,q) pa vrijedi $f(n,m) = f(i,q) + g(i,q)$ (za q iz tvrdnje teorema). Opet, jer je $f(n,m)$ duljina najkraćeg puta, a $f(i,q') + g(i,q')$ duljina nekog puta, vrijedi $f(n,m) \leq f(i,q') + g(i,q')$, $q \in J \cup \{0\}$, iz čega slijedi druga tvrdnja teorema. $\square$

Ideja algoritma za optimalno poravnanje

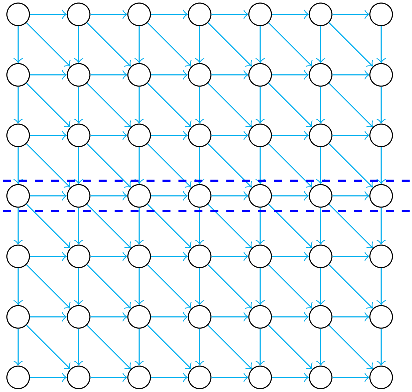
- Promatramo problem
  $A = A(x_1 \ldots x_n, y_1 \ldots y_m) = (x_1 \ldots x_n, y_1 \ldots y_m, \alpha, \delta)$



Ideja algoritma za optimalno poravnanje

- Promatramo problem
  $A = A(x_1 \dots x_n, y_1 \dots y_m) = (x_1 \dots x_n, y_1 \dots y_m, \alpha, \delta)$

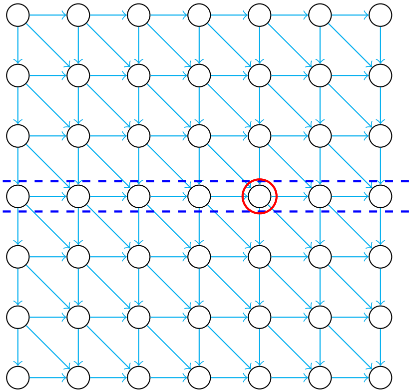
- Izaberemo redak i



Ideja algoritma za optimalno poravnanje

- Promatramo problem
  $A = A(x_1 \dots x_n, y_1 \dots y_m) = (x_1 \dots x_n, y_1 \dots y_m, \alpha, \delta)$

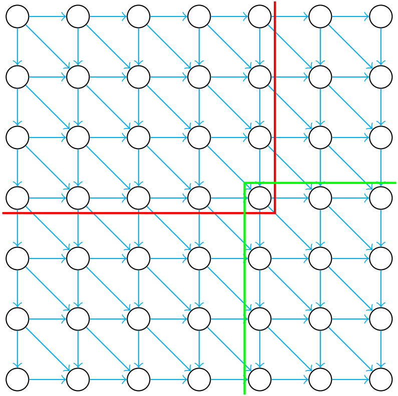
- Izaberemo redak i
- Pronađemo q iz teorema



Ideja algoritma za optimalno poravnanje

- Promatramo problem $A = A(x_1 \ldots x_n, y_1 \ldots y_m) = (x_1 \ldots x_n, y_1 \ldots y_m, \alpha, \delta)$

- Izaberemo redak i
- Pronađemo q iz teorema
- $o(n, m) = f(i, q) + g(i, q)$
- Podgraf induciran s $(i', j') \leq (i, q)$ – graf za $A_1 = A(x_1 \ldots x_i, y_1 \ldots y_q)$
- Podgraf induciran s $(i', j') \geq (i, q)$ – graf za $A_2 = A(x_{i+1} \ldots x_n, y_{q+1} \ldots y_m)$



Ideja algoritma za optimalno poravnanje (2)

- Neka je P_1 optimalno poravnanje za A_1, a P_2 optimalno poravnanje za A_2.
- Zbog veze poravnanja i najkraćeg puta u pridruženom grafu je $c(P_1) = f(i, q)$, $c(P_2) = g(i, q)$
- Definiramo
  $P_2' = P_2 + (i, q) = \{(i' + i, j' + q) \quad | \quad (i', j') \in P_2\}$
- $P_1 \cup P_2'$ je poravnanje za A i
  $c_A(P_1 \cup P_2') = c_{A_1}(P_1) + c_{A_2}(P_2) = f(i, q) + g(i, q) = o(n, m)$
  - $P_1 \cup P_2'$ je optimalno poravnanje za A
- Možemo sastaviti "podijeli pa vladaj" algoritam za računanje poravnanja
- Oznake
  - ALIGNDP – algoritam za poravnanje koji već znamo
  - ALIGNDPM – algoritam koji koristi $\Theta(m)$ memorije
  - ALIGNDPMR – "okrenuti" algoritam koji koristi $\Theta(m)$ memorije (računa g)

Algoritam za optimalno poravnanje

Algoritam

funkcija $\textsc{alignDC}(x, y)$
    ako $\textsc{len}(x) \leq 2$ ili $\textsc{len}(y) \leq 2$ **tada**
        vrati $\textsc{alignDP}(x, y)$
    inače
        $i \leftarrow \lfloor \textsc{len}(x)/2 \rfloor$
        $f \leftarrow \textsc{alignDPM}(x[:i], y)$
        $g \leftarrow \textsc{alignDPMR}(x[i+1:], y)$
        $q \leftarrow \textsc{argmin}(f + g)$
        $P_1 \leftarrow \textsc{alignDC}(x[:i], y[:q])$
        $P_2 \leftarrow \textsc{alignDC}(x[i+1:], y[q+1:])$
        vrati $P_1 \cup (P_2 + (i, q))$
    kraj
kraj

Analiza složenosti

- Prostorna složenost
  - Za $n \leq 2$ ili $m \leq 2$ $\textsc{align}\mathrm{DP}$ treba $(2+1)(m+1)$, odnosno $(n+1)(2+1)$ matricu, dakle $3(\max(n,m)+1)$ brojeva
  - U rekurzivnom slučaju, potrebno je $2(m+1)$ brojeva za $\textsc{align}\mathrm{DPM}$, a istu memoriju može koristiti i $\textsc{align}\mathrm{DPMR}$ s time da treba dodatnih $m+1$ brojeva za pamtiti potrebne vrijednosti od f
  - Nakon što izračunamo q tih $3(m+1)$ brojeva nam više ne treba pa tu memoriju mogu koristiti rekurzivni pozivi algoritma
  - Dakle, ukupno nam treba $3(max(n,m)+1) \in O(n+m)$ brojeva

Analiza složenosti

- Vremenska složenost
  - Iz algoritma se vidi da potrebno vrijeme za $T(n, m)$ zadovoljava sljedeću rekurziju

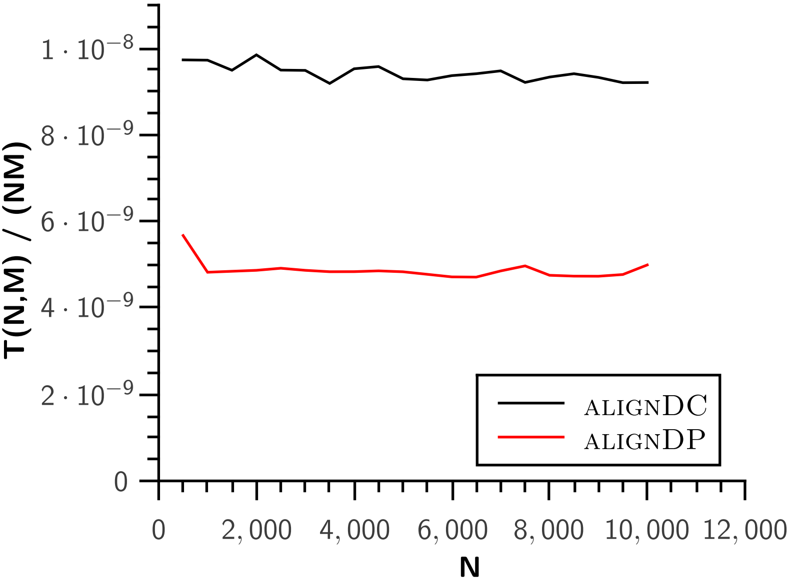
$$T(n,m) \leq cnm + T\left(\left\lfloor \frac{n}{2} \right\rfloor, q\right) + T\left(n - \left\lfloor \frac{n}{2} \right\rfloor, m - q\right), \quad n, m \geq 2$$

$$T(2,m) \leq cm$$

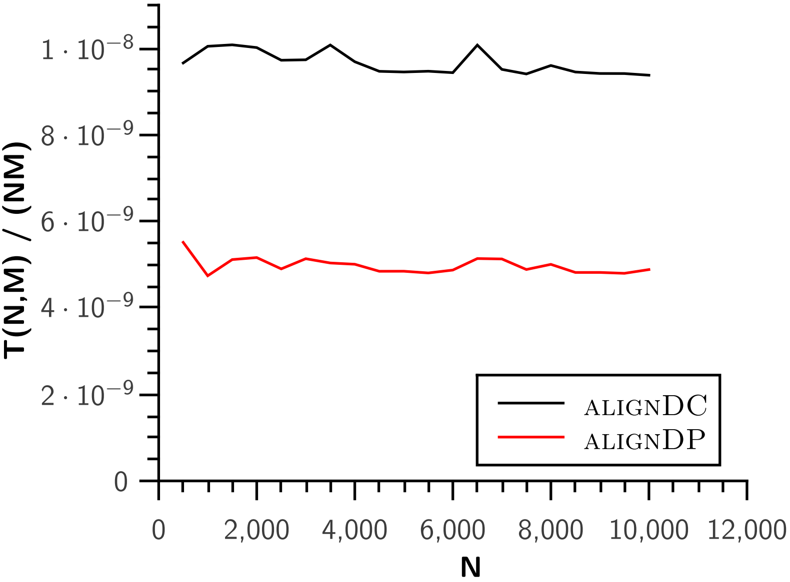
$$T(n,2) \leq cn$$

  - Konstruktivnom indukcijom po nm se može pronaći k takav da vrijedi $T(n, m) \leq knm$ pa je $T(n, m) \in \Theta(nm)$ – vrijedi za $k \geq 2c$

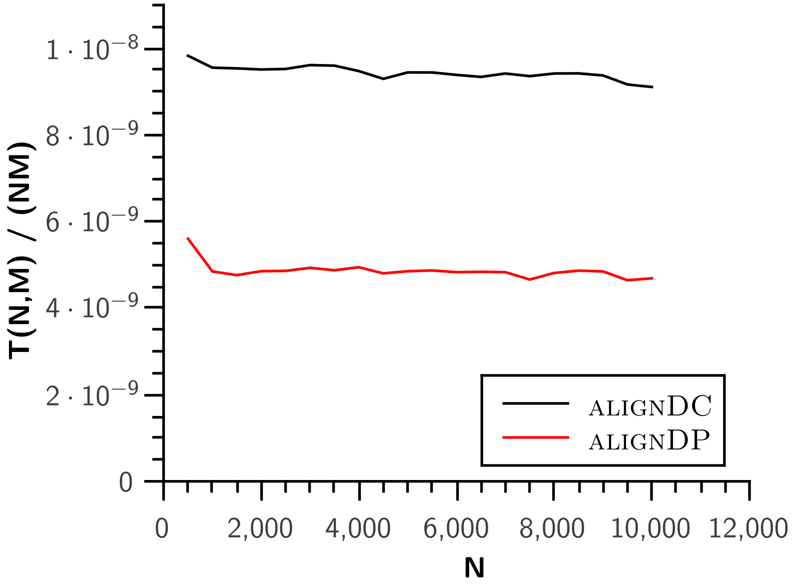
Usporedba algoritama, $M = N$



Usporedba algoritama, $M = N/2$



Usporedba algoritama, $M = N/3$



Literatura

J. Kleinberg, É. Tardos
Algorithm design
Addison-Wesley, 2006.

Sequence alignment.
Wikipedia, 12.01.2015.