

Naloga 1. - Poravnava več zaporedij

Nagrada 0 za ujemanje ter kazen 1 za zamenjavo/delekcijo/insercijo. Za razdaljo upoštevamo Levenstheimovo razdaljo = št. razlik med zaporedjema (vse kar ni ujemanje) v poravnavi.

s ₁	C	-	A	-	C	A	T	A	G	A
s ₂	T	G	A	T	T	A	-	A	G	-

ocena = -6

s ₁	-	C	A	-	C	A	T	A	G	A
s ₃	G	C	T	T	C	A	-	C	G	T

ocena = -6

s ₁	C	A	C	A	T	A	G	A
s ₄	C	G	G	-	T	A	C	A

ocena = -4

s ₂	T	G	-	A	T	T	A	-	A	G
s ₃	-	G	C	T	T	C	A	C	G	T

ocena = -7

s ₂	T	G	A	T	T	A	-	A	G
s ₄	C	G	-	G	T	A	C	A	-

ocena = -5

s ₃	G	C	T	T	C	A	C	G	T
s ₄	-	C	G	G	T	A	C	-	A

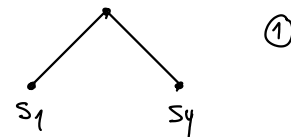
ocena = -6

Iz ocen sestavimo matriko razdalj. Upoštevamo razdalja = 1 ocena.

	s ₁	s ₂	s ₃	s ₄
s ₁	0	6	6	4
s ₂	6	0	7	5
s ₃	6	7	0	6
s ₄	4	5	6	0

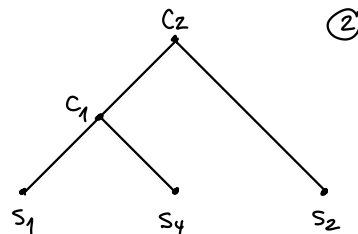
Iz matrike začnemo graditi drevo. Uporabil bom metodo UPAMA.

$$C_1 = \{s_1, s_4\}$$



$$d(C_1, s_2) = \frac{d(s_1, s_2) + d(s_4, s_2)}{|C_1|} = \frac{6 + 5}{2} = 5, \quad d(C_1, s_3) = \frac{d(s_1, s_3) + d(s_4, s_3)}{|C_1|} = \frac{6 + 6}{2} = 6$$

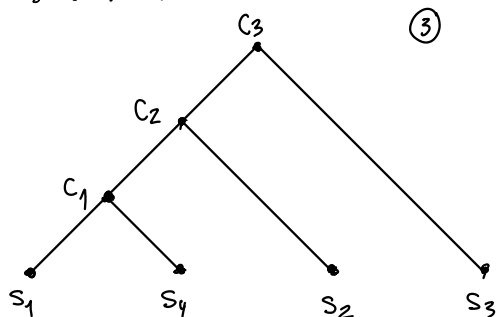
	C ₁	s ₂	s ₃
C ₁	0	5	6
s ₂	5	0	7
s ₃	6	7	0



$$C_2 = \{C_1, s_2\}$$

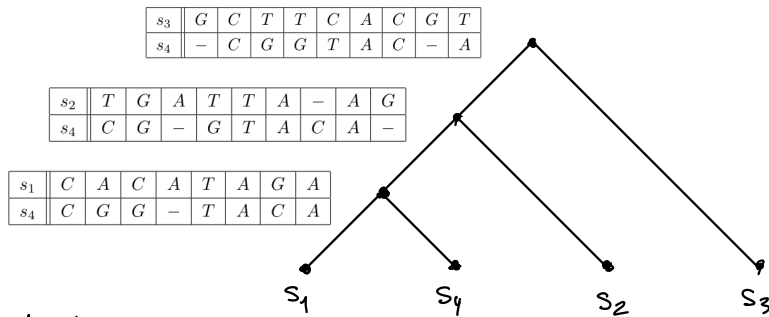
Tu lahko že opazimo, da se bo s₃ vezal s c₂, zato računanje razdalj ni več potrebno.

$$C_3 = \{C_2, s_3\}$$



Vsaki točki določimo poravnavo in sicer najboljšo poravnavo zaporedje ene skupine z zaporedjem druge skupine (oz. poddrevesa).

1. Za c₁ imamo možni samo dve zaporedji, zato c₁ dodelimo poravnavo (s₁, s₄).
2. C₂ dodelimo poravnavo, ki zadostuje $\min\{C_1, s_2\} = \min\{d(s_1, s_2), d(s_4, s_2)\} = \min\{6, 5\} \Rightarrow (s_4, s_2)$.
3. C₃ dodelimo poravnavo $\min\{C_2, s_3\}$, kjer zaporedja za C₂ obravnavamo le s₄ in s₂.
 $\min\{d(s_4, s_3), d(s_2, s_3)\} = \min\{6, 7\} \Rightarrow (s_4, s_3)$



Zaporedja ravnamo od vrha drevesa navzdol. Začnemo s poravnavo (s_3, s_4) in ji dodamo zaporedje s_2 , tako da dodamo vrzeli glede na to kako se s_2 poravnava z s_2 in s_3 :

		od (s_4, s_2)					od (s_4, s_2)				
s_3	G	C	T	-	T	C	A	C	G	T	-
s_4	-	C	G	-	G	T	A	C	-	A	-
s_2	-	T	G	A	T	T	A	-	-	A	G
		od (s_4, s_3)					od (s_4, s_3)				

Vsem dodamo še s_1 in spet dodamo vrzeli glede na to kako se s_4 poravnava z s_1 ter s_2 in s_3 :

Končna poravnava:

s_3	G	C	T	-	T	C	A	C	G	T	-
s_4	-	C	G	-	G	T	A	C	-	A	-
s_2	-	T	G	A	T	T	A	-	-	A	G
s_1	-	C	A	C	A	T	A	G	-	A	-
		od (s_4, s_3)									

