

6. (2 boda)

Nacrtajte sva moguća neukorijenjena filogenetska stabla za sljedove: S_1 , S_2 , S_3 .

Odgovor:

7. (2 boda)

Kako se zove prvi samostojni indeks (eng. *self-index*)? Kolika je prostorna složenost takvoga indeksa u odnosu na tekst nad kojim je indeks izgrađen: sublinearna, linearna ili supralinearna?

Odgovor:

8. (2 boda)

Nabrojite dvije metode za izgradnju filogenetskih stabala temeljene na udaljenosti.

Odgovor:

9. (2 boda)

Objasnite pojmove preklapanja i udaljenosti između dva niza (možete i nacrtati).

Odgovor:

10. (2 boda)

Navedite uvjete za postojanje Eulerove staze u grafu.

Odgovor:

Rješenja:

1. (5 bodova)

- a) Odredite Burrows-Wheelerovu transformaciju niza $S = \text{ACGCGAT\$}$. Pretpostavite da je znak za kraj niza abecedno manji od znakova abecede nad kojima je izgrađen ulazni niz.
- b) Skicirajte traženje niza $P = \text{GCG}$ u nizu $\text{CAGAGAC\$}$ korištenjem LF-mapiranja.

Rješenje:

a)

1	2	3	4	5	6	7	8
A	C	G	C	G	A	T	\$
\$	A	C	G	C	G	A	T
T	\$	A	C	G	C	G	A
A	T	\$	A	C	G	C	G
G	A	T	\$	A	C	G	C
C	G	A	T	\$	A	C	G
G	C	G	A	T	\$	A	C
C	G	C	G	A	T	\$	A

1	2	3	4	5	6	7	8
\$	A	C	G	C	G	A	T
A	C	G	C	G	A	T	\$
A	T	\$	A	C	G	C	G
C	G	A	T	\$	A	C	G
C	G	C	G	A	T	\$	A
G	A	T	\$	A	C	G	C
G	C	G	A	T	\$	A	C
T	\$	A	C	G	C	G	A

$\text{BWT}(S) = \text{T\$GGACCA}$

b) $\text{BWT}(\text{CAGAGAC\$}) = \text{CGGCA\$AA}$

$P = \text{GC}\mathbf{G}$

1	2	3	4	5	6	7	(8)
(F)							(L)
\$	C	A	G	A	G	A	C
A	C	\$	C	A	G	A	G
A	G	A	C	\$	C	A	G
A	G	A	G	A	C	\$	C
C	\$	C	A	G	A	G	A
C	A	G	A	G	A	C	\$
G	A	C	\$	C	A	G	A
G	A	G	A	C	\$	C	A

Niz P nije podniz niza $\text{CAGAGAC\$}$.

2. (5 bodova)

Izračunajte entropije niza S : $H_0(S)$ i $H_1(S)$ za zadani niz $S = \text{CTTT}$.

Rješenje:

$S = \text{CTTT}$, $n = 4$, $n_c = 1$, $n_T = 3$

$$H_0(S) = -\left(\frac{n_c}{n} \log \frac{n_c}{n} + \frac{n_T}{n} \log \frac{n_T}{n}\right) = -\left(\frac{1}{4} \log \frac{1}{4} + \frac{3}{4} \log \frac{3}{4}\right) = 0.811$$

$$\text{con}_1 = C, S^{\text{con}1} = T$$

$$\text{con}_2 = T, S^{\text{con}2} = TT$$

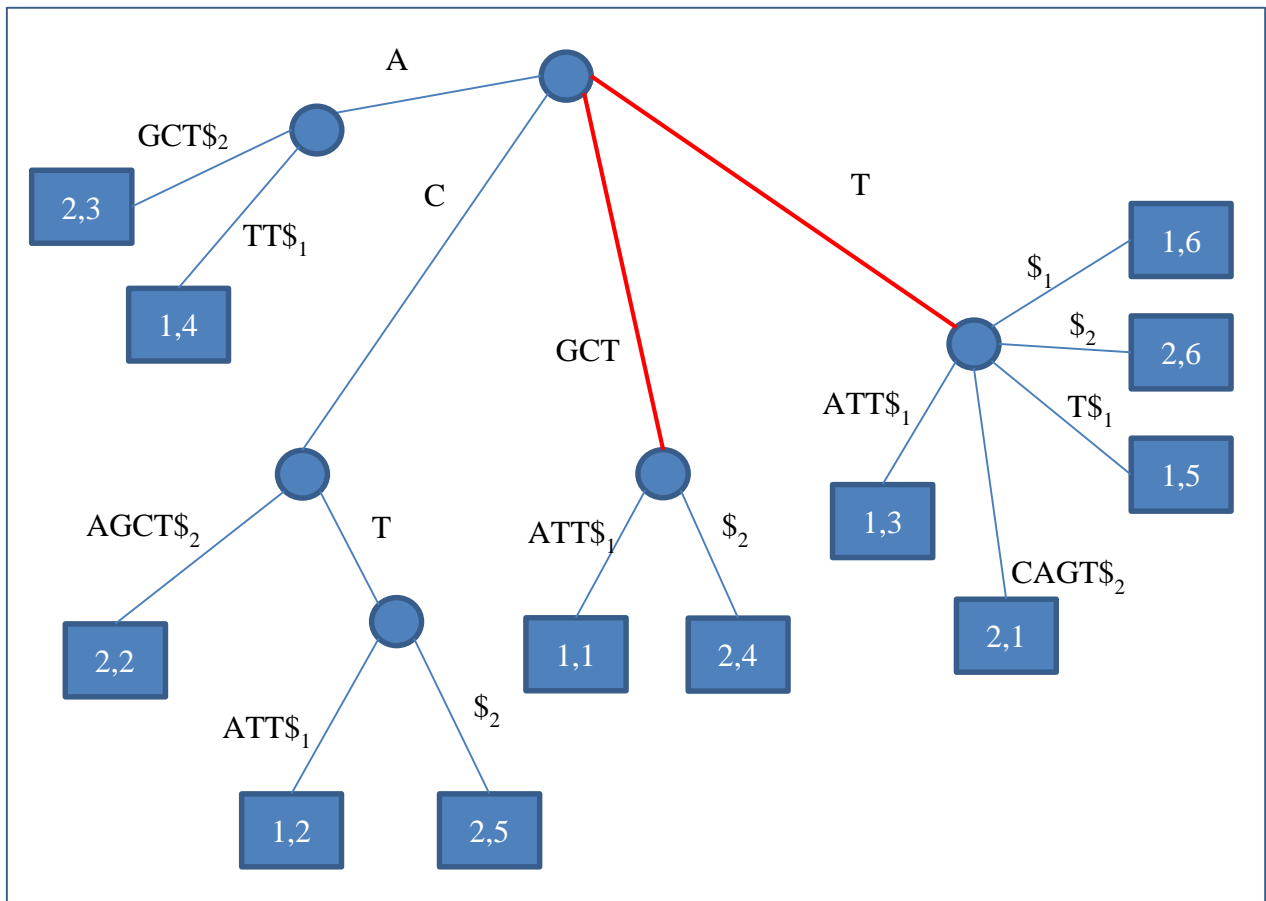
$$H_0(S^{\text{con}1}) = H_0(S^{\text{con}2}) = 0$$

$$H_1(S) = \frac{1}{n} \sum_{\text{con} \in \Sigma^1} |S^{\text{con}}| H_0(S^{\text{con}}) = 0.25 \cdot (|S^{\text{con}1}| \cdot H_0(S^{\text{con}1}) + |S^{\text{con}2}| \cdot H_0(S^{\text{con}2})) = 0$$

3. (5 bodova)

Za očitavanja $s=\text{GCTATT}$ i $t=\text{TCAGCT}$ nađite maksimalna preklapanja koristeći poopćeno sufiksno stablo. Konstruirati stablo (ne treba ucrtati sufiksne veze), ucrtati i ispisati preklapanje u obliku poravnanja (naznačene praznine).

Rješenje:



Poravnanja:

GCTATT-----
-----TCAGCT

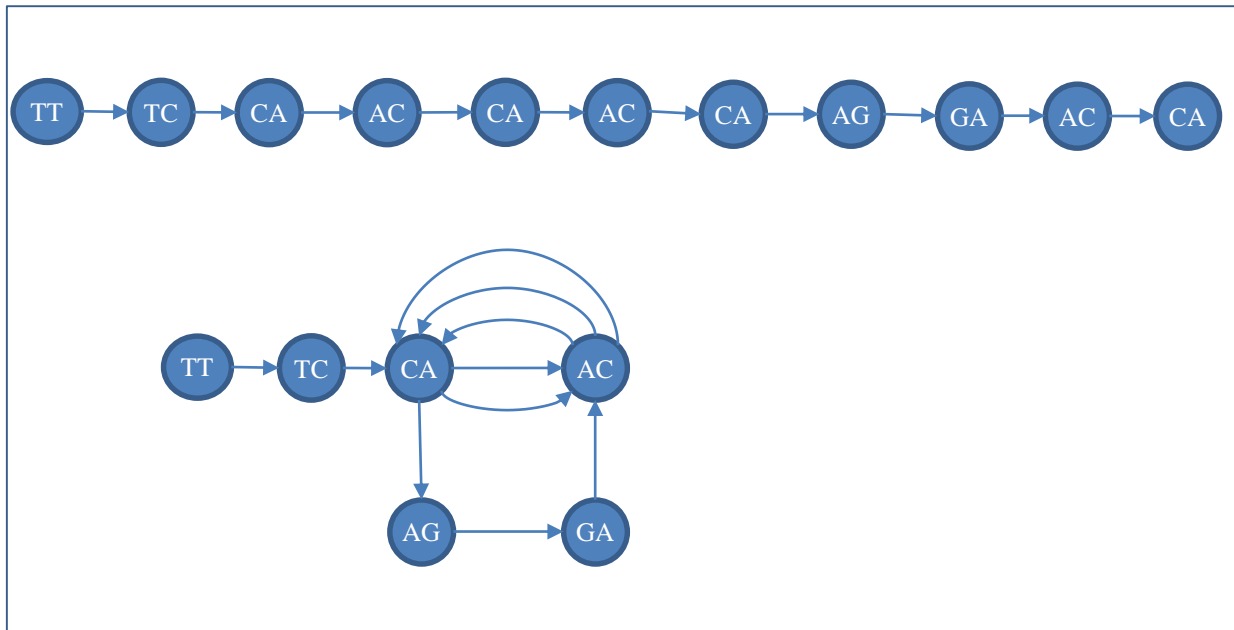
TCAGCT---
---GCTATT

4. (5 bodova)

Za niz $s = \text{TTCACACAGACA}$ napraviti očitavanja koristeći k -torke (k -torke predstavljaju niz uzastopnih nukleotida, a počinju sa svakim nukleotidom u nizu osim zadnjih $k-1$, npr. prva je TTC). $K=3$. Na osnovi očitavanja nacrtati pojednostavljeni de Bruijn graf i pronaći sve Eulerove staze u njemu i na osnovu njih ispisati moguće izlazne nizove.

Rješenje:

Očitavanja: {TTC, TCA, CAC, ACA, CAC, ACA, CAG, AGA, GAC, ACA}



Nizovi:

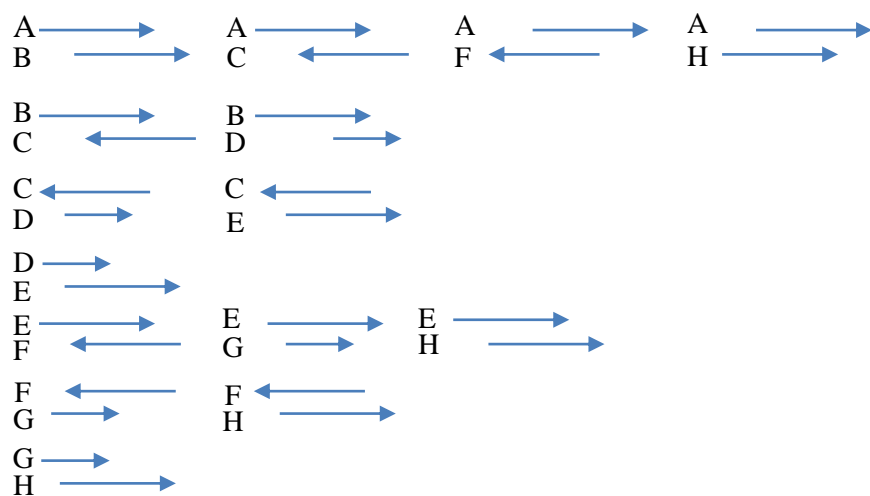
TTCAGACACACA

TTCACAGACACA

TTCACACAGACA

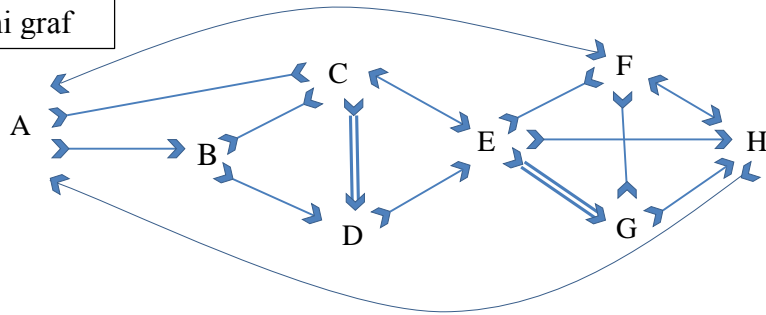
5. (5 bodova)

Za zadana preklapanja nacrtajte zajednički graf preklapanja te isti pojednostavite koristeći OLC pristup. Označite dobivene blokove.

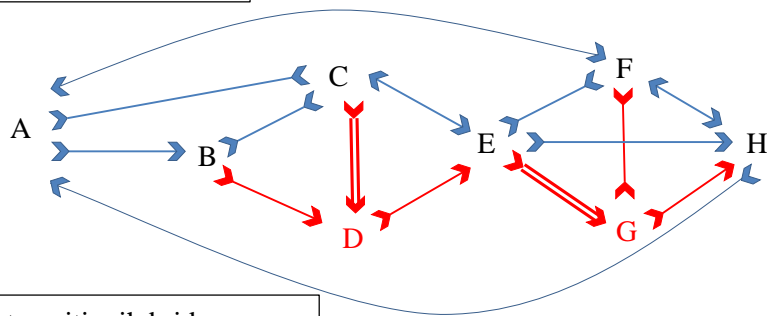


Rješenje:

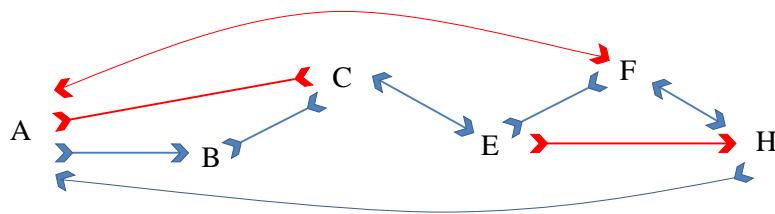
Početni graf



Izbacivanje sadržanih očitanja



Izbacivanje tranzitivnih bridova



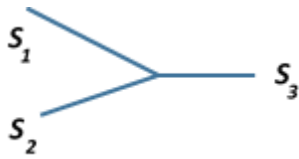
Konačni graf sadrži jedan blok



6. (2 boda)

Nacrtajte sva moguća neukorijenjena filogenetska stabla za sljedove: S_1 , S_2 , S_3 .

Odgovor:



7. (2 boda)

Kako se zove prvi samostojni indeks (eng. *self-index*)? Kolika je prostorna složenost takvoga indeksa u odnosu na tekst nad kojim je indeks izgrađen: sublinearna, linearna ili supralinearna?

Odgovor:

Prvi samostojni indeks je FM-indeks (Ferragina i Manzini, 2000). Prostorna složenost je sublinearna u odnosu na tekst nad kojim je indeks izgrađen.

8. (2 boda)

Nabrojite dvije metode za izgradnju filogenetskih stabala temeljene na udaljenosti.

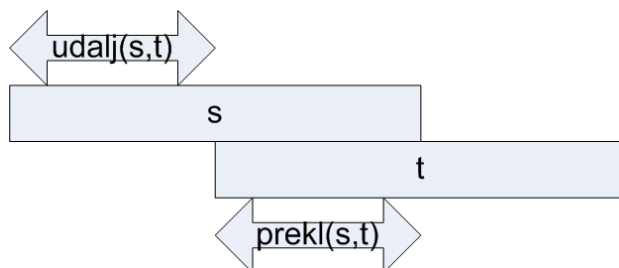
Odgovor:

UPGMA, metoda povezivanja susjeda.

9. (2 boda)

Objasnite pojmove preklapanja i udaljenosti između dva niza.

Odgovor:



10. (2 boda)

Navedite uvjete za postojanje Eulerove staze u grafu.

Odgovor:

Ulazni stupanj vrha je broj bridova koji ulaze u njega
Izlazni stupanj vrha je broj bridova koji iz njega izlaze

Da bi povezani usmjereni graf imao Eulerovu stazu potrebno je ispuniti slijedeće nužne i dovoljne uvjete:

- Najviše jedan vrh ima $(\text{izlazni stupanj} - \text{ulazni stupanj}) = 1$
- Najviše jedan vrh ima $(\text{izlazni stupanj} - \text{ulazni stupanj}) = -1$
- Svi ostali vrhovi imaju $(\text{izlazni stupanj} - \text{ulazni stupanj}) = 0$