

Uvod u Bioinformatiku
Industrija 4.0
Školska godina 2022/2023

Uputstvo:

Na Desktop-uu se nalazi folder sa nazivom BIO.SEP1 u kome se nalaze nepotpuni kodovi rešenja zadataka koje treba dopuniti kako bi traženi algoritmi bili implementirani. Folder preimenovati u format bio_ispitniRok_Prezime_Ime_BrIndeksa (npr. bio_sept1_Peric_Pera_4005-2021) i sve kodove čuvati u preimenovanom folderu dopunjujući nepotpune kodove.

1. [12.5p] Napisati kod *Frequent Words With Mismatches* algoritma koji pronalazi skup najzastupljenijih podniski dužine k u datoj sekvenci sa najviše d dozvoljenih odstupanja.

Primer:

```
text = 'CATAAATTTTCGTATGTATCAAAATTTTGTACTATCACATAAATTTTCGTATGTATCAAAATTTTGTACTATCA'
k = 6
d = 1
```

Izlaz: ['AAATTT', 'AATTTT', 'AAAATT']

2. [12.5p] Napisati kod algoritma *Smith-Waterman* koji pronalazi optimalno lokalno poravnanje između dve sekvence, pri čemu je skor uparivanja istih karaktera jednak 1, skor uparivanja različitih karaktera 0, a poravnanje karaktera sa prazninom -1.

Primer:

```
seq1 = 'OIUGNOISGRVISDIODSHFGMLSVEODRGMSOVFOGHLSDGMOVIDFHGS'
seq2 = 'DSHFGMLSVEOD'
```

Izlaz: Score: 0
Alignment:
'DSHFGMLSVEOD'
'DSH-GMLS-EOD'

3. [12.5p] Napisati kod *Greedy Sorting* algoritma za pronalaženje najmanjeg broja koraka kako bi se sinteni blokovi doveli u sortirani poredak pohlepnim pristupom.

Primer:

P = [+1, -7, +6, -10, +9, -8, +2, -11, -3, +5, +4]

Izlaz: 11

4. [12.5p] Napisati kod *UPGMA* algoritma koji konstruiše ultrametrično evolutivno stablo za matrici rastojanja D dimenzije $n \times n$.

Primer:

```
D = [[0, 3, 4, 3],
      [3, 0, 4, 5],
      [4, 4, 0, 2],
      [3, 5, 2, 0]]
```

n = 4

Izlaz: Root = [0, 3, 1, 2]:9.333333333333334
0
Neighbors: []
1
Neighbors: []
2
Neighbors: []
3
Neighbors: []
[1, 2]:6.0
Neighbors: [(' [1]:0', 6.0), (' [2]:0', 6.0)]
[3, 1, 2]:6.5
Neighbors: [(' [3]:0', 6.5), (' [1, 2]:6.0', 0.5)]
[0, 3, 1, 2]:9.333333333333334
Neighbors: [(' [0]:0', 9.333333333333334), (' [3, 1, 2]:6.5', 2.8333333333333334)]