

Proszę wpisywać odpowiedzi w ramkach

Zadanie 1 (3 pkt)

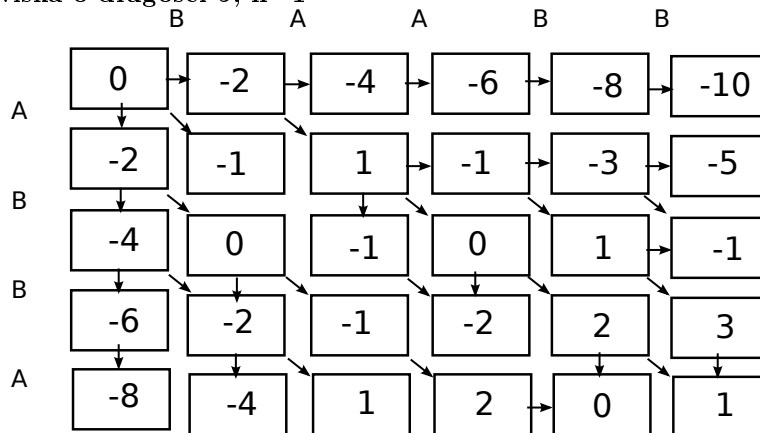
Badamy podobieństwo globalne algorytmem Needlemana-Wunscha dla sekwencji BAABB i ABBA. Stosujemy liniową karę za przerwę $d = -2$. Macierz podobieństwa pokazano obok, n jest liczbą liter Twojego nazwiska podzieloną na 3, dzielenie całkowite, bez reszty.

	A	B
A	$3 * n$	$-n$
B	$-n$	$2 * n$

Uzupełnij macierz podobieństwa, macierz pomocniczą. Podaj najlepsze dopasowanie, jeżeli jest kilka, podaj jedno z nich.

Rozwiązanie dla nazwiska o długości 5, $n=1$

	A	B
A	3	-1
B	-1	2

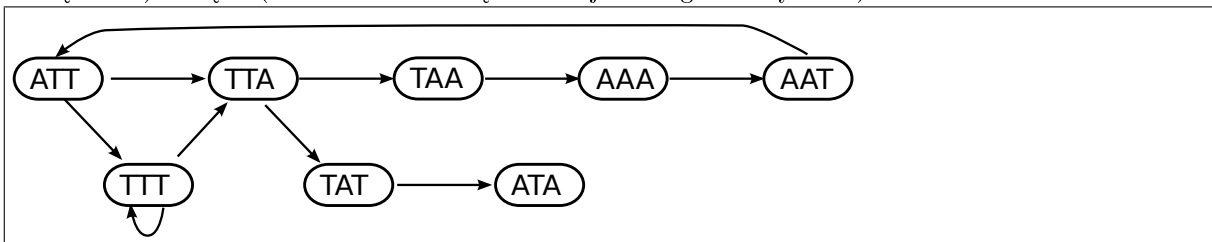


Rozwiązanie:

BAABB
-ABBA

Zadanie 2 (3 pkt)

Seqwencje TAAA, TATA, AAATTTT, ATTAA i TTTAT użyto do asemblacji pod-grafem de Bruijna (bez wag w krawędziach) 4 rzędu (w wierzchołkach są sekwencje o długości 3 symboli).



Odpowiedz na pytanie dotyczące wyniku asemblacji, wpisując a, b albo c:

- a) wynik nie istnieje,
- b) istnieje dokładnie jedna sekwencja wyjściowa,
- c) istnieje więcej niż jedna sekwencja wyjściowa.

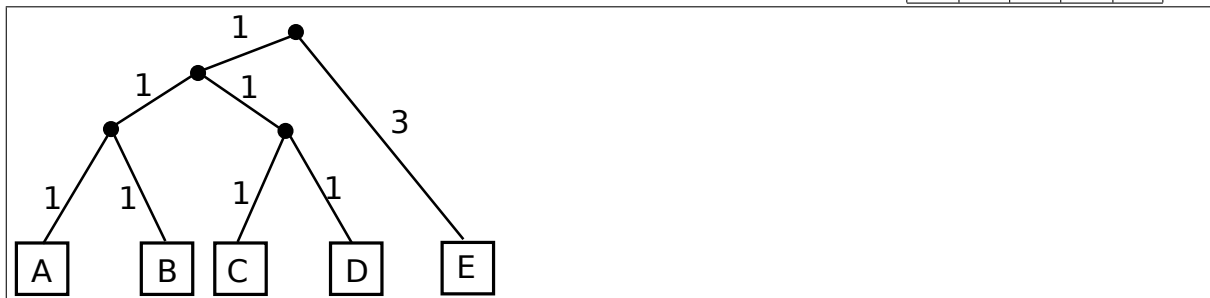
Jeżeli istnieje sekwencja wyjściowa (jedna lub kilka) jest nią:

ATTAAATTTTATAT

Zadanie 3 (3 pkt)

Używamy metody grupowania parami ze średnią arytmetyczną (UPGMA) dla 5 taksonów (A, B, C, D, E). Macierz odległości podano obok.

	B	C	D	E
A	2	3	5	5
B		5	3	7
C			2	5
D				7



Zadanie 4 (3 pkt)

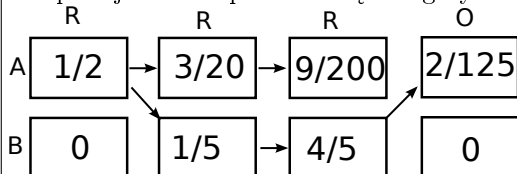
Posługujemy się monetami A i B, obserwując sekwencje rzutów *RRRO* (orły i reszki). Podaj najbardziej prawdopodobną sekwencję użytych monet, zakładając, że to doświadczenie może być opisane ukrytym modelem Markowa pokazanym niżej.

$Q = \{A, B\},$
 $V = \{O, R\},$
 $P_A = 1, P_B = 0.$

	A	B
A	$\frac{3}{5}$	$\frac{2}{5}$
B	$\frac{1}{5}$	$\frac{4}{5}$

	O	R
A	$\frac{1}{2}$	$\frac{1}{2}$
B	0	1

Uzupełnij macierz pomocniczą w algorytmie Viterbiego



Podaj najbardziej prawdopodobną sekwencję użytych monet:

A	B	B	A
---	---	---	---

Pytanie (2 pkt)

Jakie zagadnienie / zagadnienia warto byłoby dodać do MBI

Notatki lub uwagi do Roberta Nowaka:

Zadanie 5 (3 pkt)

Poniżej podano fragment pliku SAM (kolumny POS CIGAR SEQ).
oraz odpowiadający fragment sekwencji kodującej

1	6M	tatact
2	6M	atacta
3	6M	tattag
4	6M	attagc

```
>chrZ:1-15T  
atactagcttgcat
```

Opis standardu CIGAR oraz tablica kodonów jest na kolejnej stronie.

1. Określ sekwencję aminokwasów, które są kodowane przez tę sekwencję zakładając, że na pozycji 1 zaczyna się nowy kodon

2. Czy dane z pliku SAM pozwalają wnioskować o wystąpieniu wariantu(ów)?

Jeżeli tak:

- a) określ typ wariantu(ów) (SNV / Indel ?),
- b) jaki jest jego (ich) efekt na białko,
- c) czy jest (są) to wariant(y) hetero czy homo-zygotyczny(e)?

3. Zaproponuj modyfikację pliku SAM tak aby znalazły się dowody na wystąpienie jedno-nukleotydowej delecji heterozygotycznej.

Notatki lub uwagi do Tomasza Gambina:

