

Wstęp do bioinformatyki

Nr ćwiczenia: 5

Temat ćwiczenia: Dendrogram UPGMA

Nazwisko i Imię prowadzącego kurs: dr inż. Witold Dyrka

Wykonawcy:	
Imię i Nazwisko Nr indeksu, wydział	Edyta Krukowska 217097, WPPT
Termin zajęć: dzień tygodnia, godzina	Piątek 11.15
Data oddania sprawozdania	06.06.2019

Repozytorium: <https://github.com/Edie1995/Bioinformatyka/tree/zadanie5>

1. Prezentacja działania programu:

```
>seq1
AAACGT---
>seq2
---CGT---
>seq3
---CGTAAA
>seq4
---CAT---
>seq5
---CCT-TA
```

```
((seq4:0.5,seq2:0.5):1.5,seq5:2):0.5,seq3:2.5)
```

2. Oszacowanie złożoności czasowej obliczeniowej i pamięciowej kodu poszczególnych funkcji i całego programu:

➤ Czasowa:

createMatrix:

3 operacje if + 4 przypisania wartości * $(m)(n) = O(mn)$;

searchingWay:

Przyjmując pesymistyczny wariant szukania maksimum $= (m)*(n)*(n)$

Funkcja while w pesymistycznej wersji obejmie $(m)*(n)*(3 \text{ porównania i } 6 \text{ przypisań wartości})$

Podsumowując, w negatywnym wariancie cała funkcja ($c > 0$) $c*(m)(n)$ więc jest ona maksymalnie rzędu $O(mnn)$

saveToFile:

W pesymistycznym założeniu, lecz raczej nie możliwym do osiągnięcia, że ścieżka będzie szła wzdłuż krawędzi macierzy jest to funkcja rzędu $O(m+n)$

writeSequence:

W funkcji while według założenia pesymistycznego otrzymamy $m*n*(6 \text{ porównań (maksymalnie) i } 2 \text{ przypisania})$. Jest to największy while w tej funkcji, pozostałe są rzędu $o(mn)$, czyli na pewno niższego niż $O(mn)$. Z takiego oszacowania wynika, że maksymalny tej funkcji to $O(mn)$.

divideMatrix:

W pesymistycznym założeniu 3 pętle for o długościach n , w nich dwie instrukcje warunkowe if. Funkcja maksymalnie rzędu $O(n^3)$.

childMatrix:

Pętla for w pętli for w najbardziej zagnieżdżonej pętli instrukcja if. W pesymistycznym założeniu obie pętle o długości n co daje nam szacunkową złożoność maksymalnie rzędu $O(n^2)$.

minDistance:

Pętla for o pesymistycznej długości n , do tego wewnątrz niej warunek if. Daje to maksymalnie funkcję rzędu $O(n)$.

Pozostałe funkcje wykorzystywane przez program są rzędu stałego $O(1)$ lub liniowego $O(n)$ lub $O(m)$. Na tej podstawie można oszacować, że dla całego programu jest maksymalnie rzędu $O(n^3)$.

➤ Pamięciowa

Największym obiektem stworzonym w całym programie jest macierz kosztu i jest ona rzędu $O(mnn)$, wszystkie pozostałe obiekty wykorzystywane w programie są rzędu takiego samego, lub niższego $o(mn) < O(mn)$. Możemy więc uznać, że jest to złożoność pamięciowa $c*m*n$ (gdzie $c > 0$), czyli $O(mn)$.

3. Testy

3.1.

```
>seq1
AAACGT---
>seq2
---CGT---
>seq3
---CGTAAA
>seq4
---CAT---
>seq5
---CCT-TA
```

```
stringExpected =
```

```
"(((seq4:0.5,seq2:0.5):1.5,seq5:2):0.5,seq3:2.5)"
```

```
test result: true
```

3.2.