

Wstęp do bioinformatyki

Nr ćwiczenia: 4

Temat ćwiczenia: Progresywne dopasowanie wielosekwencyjne

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 | 23.05.2019 |

Repozytorium: <https://github.com/Edie1995/Bioinformatyka/tree/zadanie3>

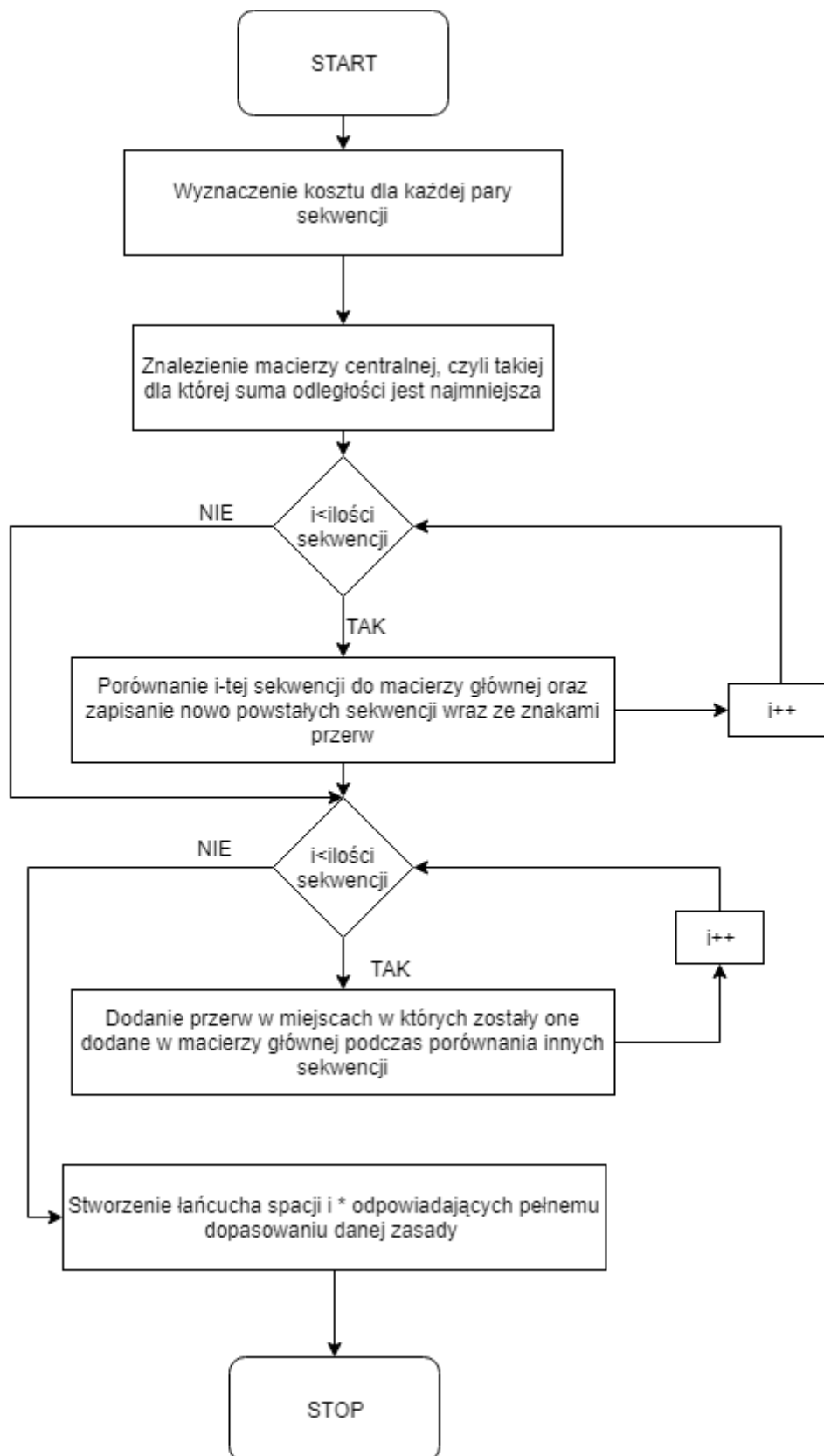
1. Prezentacja działania programu:

```
seq1    AAACG-T---      6
seq2    ---CG-T|---     3
seq3    ---CG-TAAA      6
seq4    ---C-AT---      3
          *  *
```

```
>seq1
AAACG-T---
>seq2
---CG-T---
>seq3
---CG-TAAA
>seq4
---C-AT---
```

Rysunek 1 Przykład działania programu, dla przykładowych sekwencji wczytanych z przykładowych plików

2. Schemat blokowy algorytmu dopasowania



Rysunek 2 Schemat blokowy -tworzenie dopasowania wielu sekwencji

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

➤ **Czasowa:**

addGapSign:

3 operacje if w 2 pętlach for, jedna pętla przy pesymistycznym założeniu porównywania n sekwencji o rozmiarze n, druga przy pesymistycznym założeniu długości sekwencji m, o rozmiarze m, złożoność maksymalnie rzędu $O(mn)$

addStar:

Podobnie jak w przypadku wcześniejszej funkcji, dwie pętle for, jedna zależna od długości sekwencji, druga od ich ilości, co daje nam ponownie złożoność maksymalnie rzędu $O(mn)$

searchingCenter:

Dwie pętle for obie o ilości iteracji równej ilości sekwencji, przy pesymistycznym założeniu, że ich ilość to n, cała funkcja maksymalnie rzędu $O(n^2)$

createMatrix:

Dwie główne pętle zagnieżdżone o ilości iteracji równej długości poszczególnych sekwencji. W pesymistycznym założeniu, że jedna z nich będzie miała długość n a druga m funkcja maksymalnie rzędu $O(mn)$

searchAligment:

Funkcja zawierająca pętlę o ilości iteracji równej ilości sekwencji sprawdzanych, w pesymistycznym założeniu n, dodatkowo pętla ta wywołuje funkcję zawierającą pętlę o ilości iteracji n w pętli o ilości iteracji n (zależą one od długości porównywanych sekwencji), całość maksymalnie rzędu $O(nm^2)$

toTextFile:

Zawiera pętlę główną o ilości iteracji równej długości sekwencji /60, pod nią zagnieżdżona jest pętla o ilości iteracji równej ilości sekwencji, a w niej pętla o ilości iteracji równej długości sekwencji, w pesymistycznym założeniu $n*m*n$. Funkcja maksymalnie rzędu $O(mn^2)$.

Ponieważ pozostałe funkcje nie zawierają tak złożonych operacji, można przyjąć, że program jest maksymalnie rzędu $O(mn^2)$.

➤ **Pamięciowa**

Największa ze struktur w programie ma rozmiar $n*m$, jest to struktura zawierająca n sekwencji o długości m. Struktura ta jest wykorzystywana w kilku funkcjach, jednak wywołania te nie są wysokiego rzędu. Można więc przyjąć, że złożoność pamięciowa programu wyniesie $O(mn)$

4. Porównanie przykładowych par sekwencji

| | | |
|------------|---|-----|
| JQ735466.1 | -A-GG-----TCN-AGC----CC-----GG-A-A-CC-----CTACT----T--- | 23 |
| JQ735460.1 | --GGG-----TC----CAA--CCT-----GG-----CA-----C-ACTGC--TA-- | 23 |
| JQ735459.1 | TA-GG-----C----C--AACC-T-----GG--GA-CC-----CTAC-----T-A- | 23 |
| JQ179960.1 | -A---TTAATTC-GAG-----C--TGAATTAGGTA-ATCC-TGGGTCT-CT--AAT--T | 40 |
| | * * ** * * * | |
| JQ735466.1 | GGC---GATGATCAA--ATCT--A-T--AATGTA---ATT-GTT-ACA--GCT-----CA | 61 |
| JQ735460.1 | GG-A--GATGATC-AG-AT-T---TTC-AATGTA---A-TCG-T--CA---CTGCC--CA | 61 |
| JQ735459.1 | GG--A-GATGATC-A-GATC-G-A---CAATGT-C--ATT-G-T-A-ACCGC----C-CA | 61 |
| JQ179960.1 | GG---TGATGATCAA--AT-T-TA-T--AA--TA-CTATT-G-TCACA--GC-----ACA | 78 |
| | ** ***** * ** ** * * * * * * * * | |
| JQ735466.1 | TGCC-T-TT-GTA--ATA-ATCTTCTT--T---ATA---GTA---ATAC-CC--ATTATG | 101 |

JQ735466.1 – cytochrom c świni (COI)

JQ735460 – cytochrom c kota (COI)

JQ735459.1 – cytochrom c konia (COI)

JQ179960.1 – cytochrom c żaby (COI)