

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>

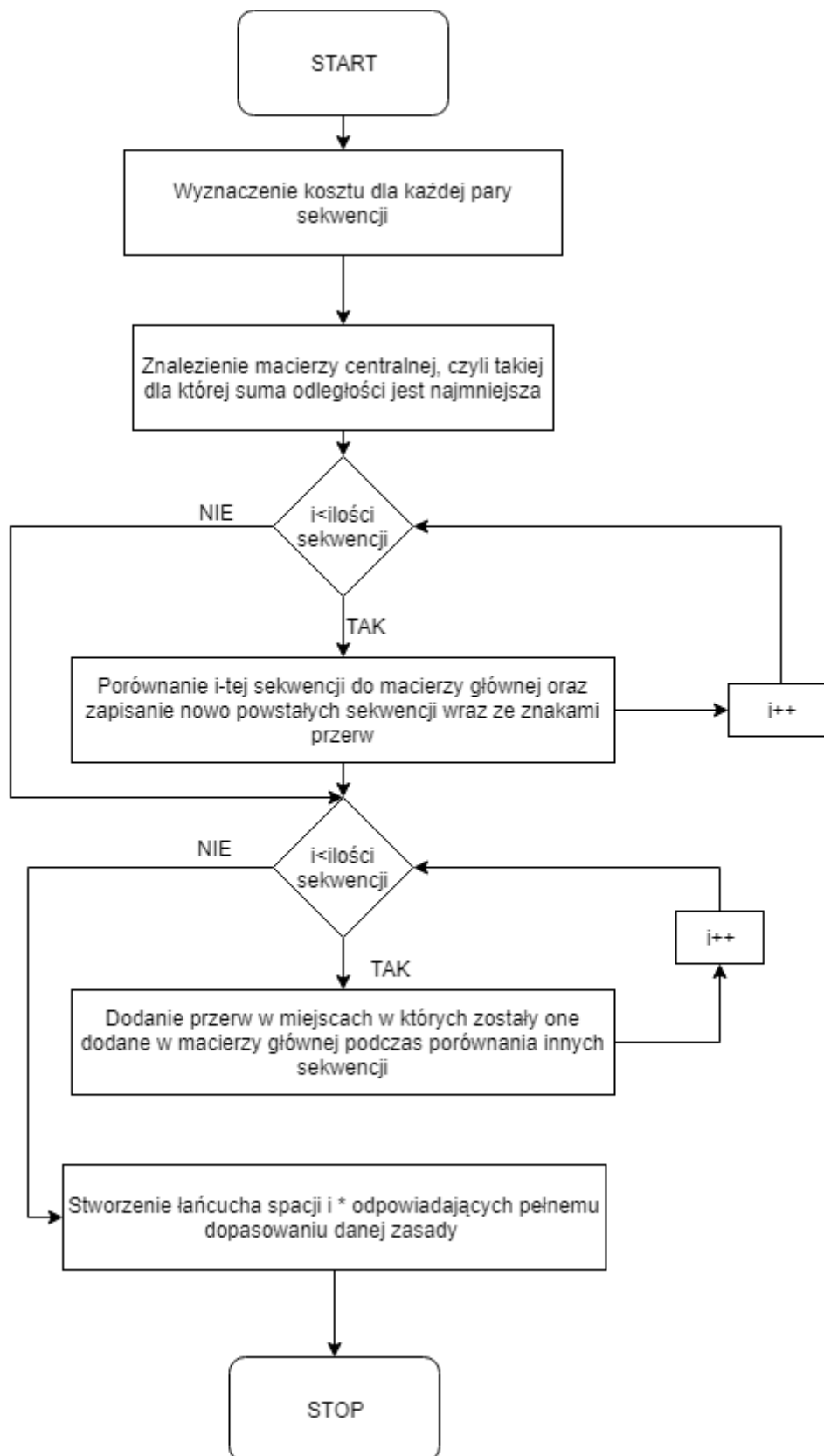
1. Prezentacja działania programu:

```
seq1    ACC---TGAC      7
seq2    -CC---T-AC     5
seq3    ACC---AG-C     6
seq4    -CCATTG-C      8
          **          *
```

```
>>seq1
ACC---TGAC
>>seq2
-CC---T-AC
>>seq3
ACC---AG-C
>>seq4
-CCATTG-C
```

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 $O(mn)$

chooseCenterSequence:

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 $O(mn)$

searchingBestAligment:

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-GGTCNAGCCC-----GG-A-A--C-----C-CTACTTGGCGATGATCAAATCT-A	40
JQ735460.1	---G-GGTCCAACCT-----GG-C-A--C-----A-CTGCTAGGAGATGATCAGAT-T-T	39
KC985188.1	AT-ATGGT-GAGCTC-----ATA--C-----G-ATA--AAGC-CT-A-AAAACCCAA	37
JQ735459.1	-T-A-GG-CCAACCT-----GG--GA--C-----C-CTACTAGGAGATGATCAGATC-GA	40
JQ179960.1	ATTA-ATTCGAGCTGAATTAGGTA-ATCCTGGGTCTCTAATTGGTGATGATCAAAT-TTA	57
	* * * * *	
JQ735466.1	T-A-ATG-TAATTGTTACAG-CTCATGCCTTTGTA--AT---AA---TCTTC-TT---T-	83
JQ735460.1	TCA-ATG-TAATCGTCACTG-CCCATGCTTTTGTGTA--AT---GA---TCTTT-TT---T-	83
KC985188.1	T-AGATA-TTATTGCT-CATACT-ATTCCTATATA--ACCGAAAGGCTCTTTTTTCCCTG	91
JQ735459.1	-CA-ATG-TCATTGTAACCG-CCCATGCATTCGTA--AT---AA-----TT--TT-CGTG	83
JQ179960.1	T-A-ATACT-ATTGTCACAG-CACATG-CTTT-TATTAT---AA---T-TT--TT--TTT	100
	* ** * ** * * * ** * ** * * ** ** *	
JQ735466.1	A-TAGTAA-T-ACCCAT-T-----ATGATTG-G-----GGGTTTTGG-TAAC--TG---A	122
JQ735460.1	A-TGGTGA-T-GCCTAT-T-----ATAATCG-GA-----GGG-TTCGG-AAAC--TG---A	122
KC985188.1	ACTAGTAACTCA-GAATCTGCAATTATTCGCAATGCGCG-GAGCATAGCAATCTAA	110

JQ735466.1 – cytochrom c świni (COI)

JQ735460 – cytochrom c kota (COI)

KC985188.1 – cytochrom c psa (COI)

JQ735459.1 – cytochrom c konia (COI)

JQ179960.1 – cytochrom c żaby (COI)