

# 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>

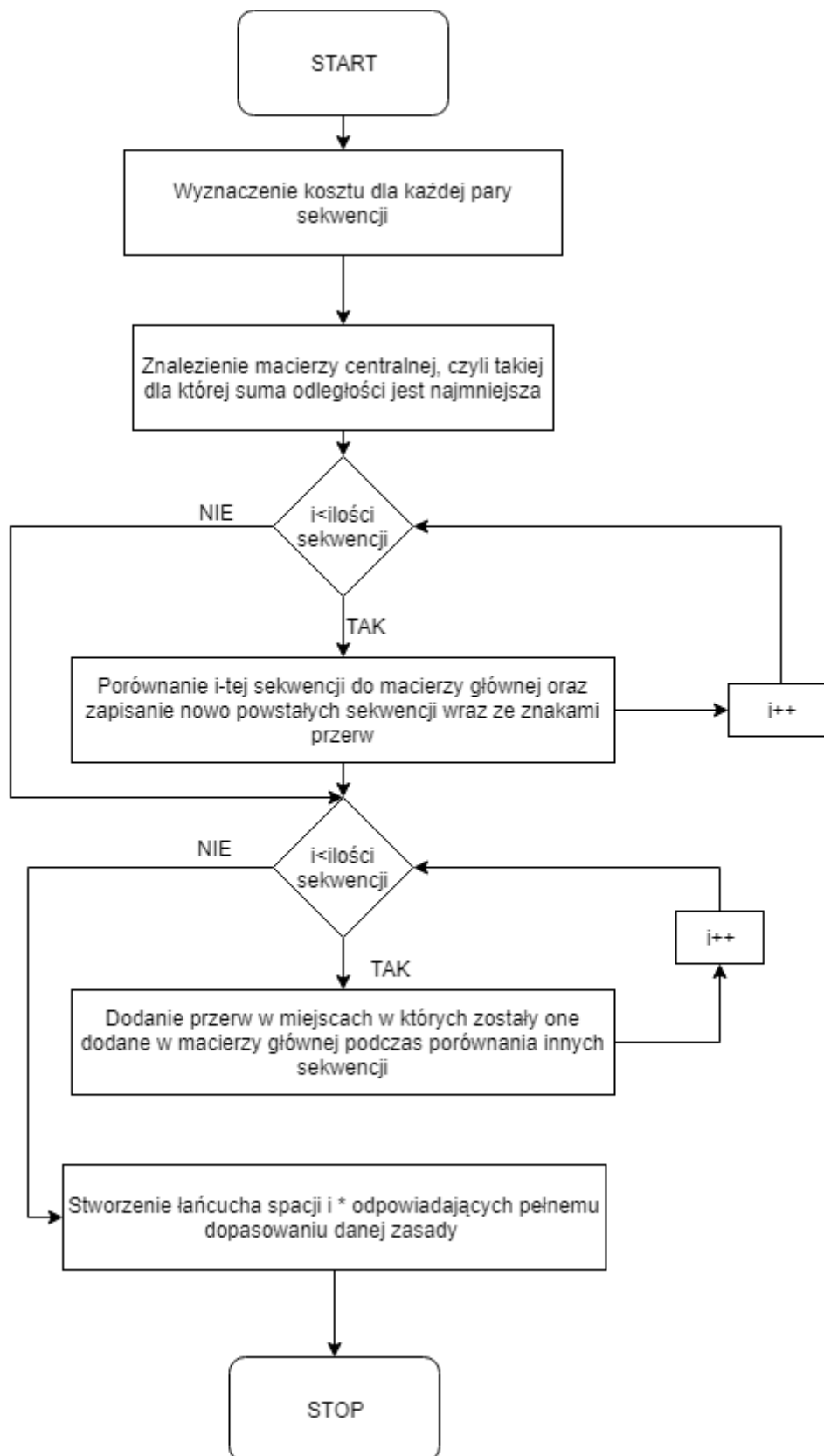
#### 1. Prezentacja działania programu:

```
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

KC591872.1	CC--GGCTCA-CTACTGGGA-----AGG-GATCAA--CT-AT--ATAATA--CAATCG	42
KC985188.1	ATATGG-TGAGCT-C---AT-----A-C-GATAAAGCCT-A--A-AA-ACCCAATAG	40
JQ735459.1	--TAGGC-CAAC--CTGGGACCCTACTAGGAGAT-GA-TCAGATCGACAAT-GTC-ATTG	51
	**  *  *  *  *          *  ***  *  *  *  *  **  *  ***  *	
KC591872.1	TTAC-A--GCCACGCC-TT--T-T-TAAT---AA---TCTTTTTTC-TCGTTA-TA-CC	85
KC985188.1	ATATTATTGCTCATACTATTCCTATATAACCGAAAGGCTCTTTTTTC--CCTGAGTA---	95
JQ735459.1	-TA--ACCGCCCATG-C-AT--T-CGTAAT---AA---T-TTTCGTGATGGTCA-TACCC	95
	**  *  ***  **          *  *  ***  **  *  ***  *          *  *  **	
KC591872.1	-TGTAAT-T-A-ATTG-GAGG-ATT-TG--GTAATT-GA-CTGG-TACC--TCTAATA-TT	130
KC985188.1	--GTAAGTGACAATGTGAGAAATTATTCCG-AATCCGGGCAGGATAAGAATGTAA-ACTT	151
JQ735459.1	ATG--A-T-A-ATCG-GAGG-ATT-CG--GAAACT-GA-TTAG-T-CC-CCCTGATAATT	140
	*  *  *  *  *  ***  ***          *  **  *          *  *  *  ***	
KC591872.1	-AGGG-GCACCG--GACAT-AGCTTTTCCTCGATTAAACAATATAAGGTTTTGG-CTACT	184
KC985188.1	CAGGATGC-CCGAAGA-ATCAG-AATAGGT-G-TT---GATAT-AGGATAGGGTCT-CC	200
JQ735459.1	GGAGGACG  TGATAT  AGCTTTTCCTCGAATTAAGCAACATAA  GCTTCGGATTACT	193