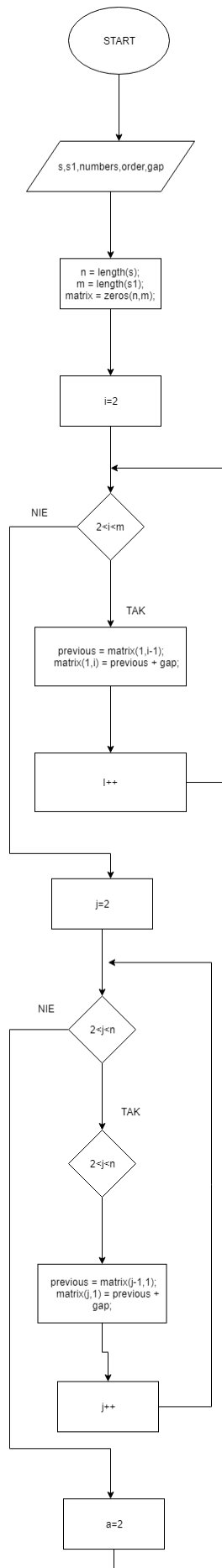


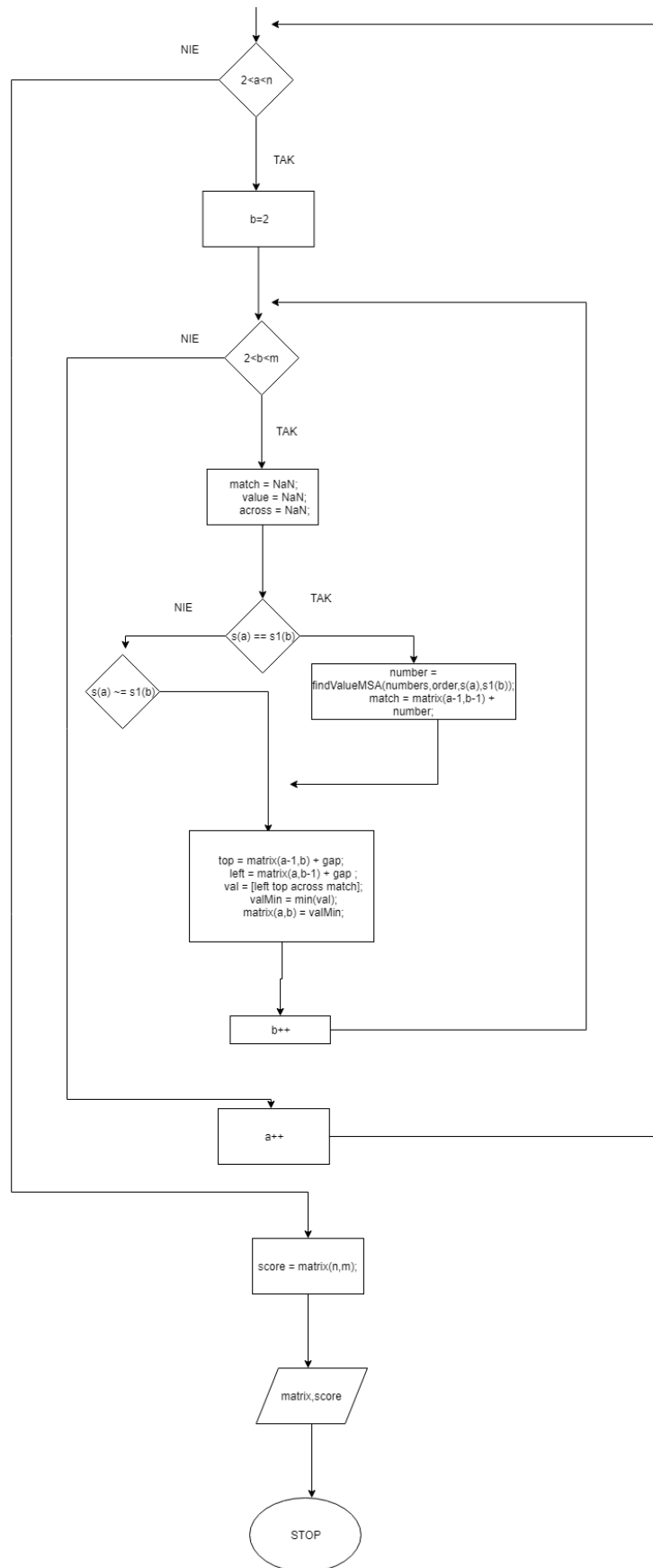
Jakub Cieplucha 230331

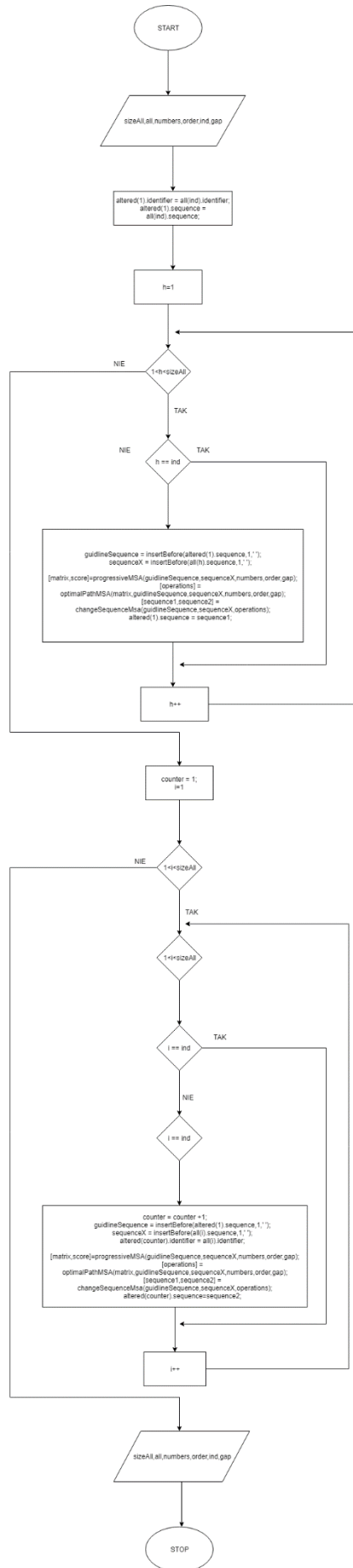
Wstęp do bioinformatyki

Zdanie 4

- 1. Schemat blokowy algorytmu dopasowania wielu sekwencji.**







2. Analiza złożoności obliczeniowej czasowej i pamięciowej.

2.1 Funkcja multipleSequenceAlignement

a) czasowa

$O(nm)$

b) pamięciowa

n - rozmiar sekwencji wejściowych

m - rozmiar sekwencji po zmianach

$S(n,m) = 4n + 4m + 2$

2.2 Funkcja progressiveMSA

a) czasowa

$O(mn^2)$

b) pamięciowa

nm – rozmiar macierzy

n + m – rozmiar sekwencji wejściowych

$S(n,m) = n + m + nm + 5$

2.3 Razem

a) Czasowa: $O(mn^2)$

b) Pamięciowa: $S(n,m) = mn + 5m + 5n + 7$

Porównanie przykładowych par sekwencji ewolucyjnie powiązanych.

```
>> zadanieMSA('filename1', 'sequence.fasta', 'filename2', 'sequence1.fasta', 'subMatrixFile', 'msa.xlsx');
Complete cost : 24
X3333333.3 G_ACA_ 4
X1111111.1 GCACAT 6
X2222222.2 TGAGA_ 5
X4444444.4 GAACT_ 5
*
```

Rysunek nr 1. Dopasowanie wielu sekwencji dla przykładu pokazanego na wykładzie.

W celu określenia skuteczności opracowanej metody dopasowania wielu sekwencji, przetestowano ją na omówionym na wykładzie wzorcu. Uzyskano niższy o 6 punktów całkowity koszt dopasowania.

```

Command Window
>> zadanieMSA('filename1','file1.fasta','filename2','file2.fasta','subMatrixFile','msa.xlsx');
Complete cost : 1422
AJ130824.1 A__TGA__CCAACATTCGTA__AA__AC 20
AJ314566.1 TCGTGAAAACCACCGTTGTTATTCAACTAC 30
Z12030.1 A__TGG__CCAACCTCCGAA__AA__AC 20
M28016.1 A__AGC__CCGA_ATGA_TA__TT__TC 18
      *   **   *   *   *
AJ130824.1 __G__C__A__T__CC__CCTACTAAAAA__ 47
AJ314566.1 AAGAACCCTAATGGCCACCTCCGAAAAACC 60
Z12030.1 __C__C__A__C__CC__CCTCCTAAAAA__ 47
M28016.1 __C__TA__T__TC__GCCTAC_ACAA__ 46
      * *   *   *** * * *
AJ130824.1 __T__TGT__TAA__C__C__ACT__C__C__ 74
AJ314566.1 CATCCTCTCCTAAAAATCGCTAATGACGCA 90
Z12030.1 __T__TGC__AAA__C__G__ACG__C__A__ 74
M28016.1 __T__TCTCCGATCCGTC__C__CTAACAAA 83
      * *   *   *   *
AJ130824.1 CTAATCGACCTTCCCGCCCCCTCAAATATC 120
AJ314566.1 CTAGTCGACCTCCAGCACCTCTAACATT 120
Z12030.1 CTAGTTGATCTCCAGCTCCTTCAAACATT 120
M28016.1 CTAGGAGGCGTCCTTGCCCTATTAC_TATC 119
      *** *   * * * * *
AJ130824.1 TCTGCCTGATGAAACTTCGGCTCTCTATTG 150
AJ314566.1 TCAGTCTGATGAAACTTTGGCTCACTCCTA 150
Z12030.1 TCTGTTTGATGAAACTTTGGCTCCCTGCTA 150
M28016.1 CAT__CCTCATCCTAGCAATAATCCCCATCC 149
      * **   *   ** *
AJ130824.1 GGATTATG__CCTAATAATCCAG_ATCCTAA 178
AJ314566.1 GGCCTATG__TTTAGCCACCCAA__ATTCTTA 178
Z12030.1 GGGCTCTG__TCTAGCTGCCCAA__ATCCTGA 178
M28016.1 TCCATATATCCAAACAACAAGCATAATAT 180
      * *   *   *   *
AJ130824.1 CTGGCTTATTCTTAGCCATACACTACACAT 210
AJ314566.1 CCGGGCTCTTCTTAGCCATACACTATACCT 210
Z12030.1 CAGGCCTCTTCTTAGCCATACATTACACCT 210
M28016.1 TTCGCCCA__CTAAGCCAATCACTTTAT__T 207
      *   *   ***** * * * *
AJ130824.1 CAGACACAGCAACAGCATTCTCCTCAGTTA 240
AJ314566.1 CCGACATTTCAACAGCTTTTTCCTCTGTCT 240
Z12030.1 CCGATATCGCCACCGCCTTTTCCTCCGTTG 240
M28016.1 GACTCCTAGCCGCAG__AC__CTCCTCATTCT 238
      *   * *   ***** *

```

```

          * * *      * * * * *
AJ130824.1 CACATATTGCCGAG_ACGTAAACTACGGC 269
AJ314566.1 GCCACATTGCCGAG_ATGTTAGTTACGGC 269
Z12030.1   CCCACATCTGCCGTG_ATGTTAATTACGGC 269
M28016.1   AACCTGAAT_CGGAGGACAACCAGTAAGCT 269
          * * * * *      * * *
AJ130824.1 TGACTIONTCGTTACTTACACGCCAATGGA 300
AJ314566.1 TGACTIONTCGAAATATCCACGCCAACGGG 300
Z12030.1   TGACTIONTCGAAACATGCACGCTAACGGC 300
M28016.1   ACCCTTTTACCATCATTGGACAAGTA_GCA 299
          * * *      * * *      * *
AJ130824.1 GCATCAATATTCTTTATTTGCTTATATATA 330
AJ314566.1 GCATCTTTCTTTTTATTTGCATTATATATA 330
Z12030.1   GCATCTTTCTTTTCATTTCATTATATCTC 330
M28016.1   TCCGTACTATACTTCACAACAATCCTAATC 330
          * * * * *      * *
AJ130824.1 CATGTAGGCCGTGGAATCTATTACGGCTCA 360
AJ314566.1 CATATCGCCCGAGGACTTTATTATGGCTCT 360
Z12030.1   CACATCGCCCGAGGCTTGTACTACGGCTCC 360
M28016.1   CTAATACCAACTATC_TCCCTAATTGAAAA 359
          * *      * * *
AJ130824.1 TATACTTACCTAGAAA_CCTGAAACATTGG 389
AJ314566.1 TACCTCTACAAAGAAA_CCTGAAATATTGG 389
Z12030.1   TACCTCTACAAAGAAA_CCTGAAACATTGG 389
M28016.1   CAAAATACTCAAATGGGCCTGAAACATTGG 390
          * *      * * * * * * *
AJ130824.1 CATTATTCTATTA_TTCGCAGTTATGGCTA 419
AJ314566.1 GGTGGTACTTCTACTTCTCACT_ATAATAA 419
Z12030.1   AGTAATTCTCCT___C_CTTTAA__CTA 413
M28016.1   CATTATTCTATTA_TTCGCAGTTATGGCTA 419
          * * * * *      * * * * *
AJ130824.1 CAGCATTTCATAGGCTATGTCCTCCCATGAG 450
AJ314566.1 CCGCCTTTGTAGGCTACGTCCTCCCATGAG 450
Z12030.1   T_G_ATA_ACAG_CT_T_T__TG__TG_G 438
M28016.1   CAGCATTTCATAGGCTATGTCCTCCCATGAG 450
          * *      * * * * * * *
AJ130824.1 GACAAATATCA 461
AJ314566.1 GACAAATATCA 461
Z12030.1   G_CTACG_TC_ 458
M28016.1   GACAAATATCA 461
          * * *      * *

```

Rysunek nr 2. Dopasowanie 4 sekwencji ewolucyjnie powiązanych.