Natalia Niewdzięczna 149514

Wadim Sokołowski 149557

Mikołaj Grochowski 149435

**Elementy Bioinformatyki**

Temat 6. Wykrywanie motywu zadanego w postaci wyrażenia regularnego PROSITE w sekwencji. Typ A.

**1. Opis problemu**

Naszym zadaniem było utworzenie programu, który na wejściu otrzymuje sekwencję aminokwasów oraz wyrażenie regularne zapisane w notacji PROSITE. Na wyjściu program zwraca najdłuższy możliwy podzbiór wejściowej sekwencji, który odpowiada zadanemu wzorcowi.

Notacja PROSITE:

* – - separator oddzielający kolejne elementy składowe wzorca
* G - litera oznaczająca konkrety aminokwas (glicyna)
* x - dowolny aminokwas
* […] - jeden aminokwas ze zbioru zdefiniowanego w nawiasach kwadratowych
* {…} - dowolny aminokwas spoza zbioru zdefiniowanego w „ostrych” nawiasach
* e(i) - powtórzenie elementu e dokładnie i razy
* e(i,j) - powtórzenie elementu e dokładnie k razy, gdzie k>=i oraz k<=j

**2. Opis programu**

Program napisany został w języku Java, przy użyciu standardowych bibliotek. Zawiera on 3 klasy:

* Main – definiuje wyrażenie regularne oraz sekwencje testowe
* AASequence – klasa ta definiuje główną logikę obsługi wejściowego wzorca zapisanego w notacji PROSITE
* Matcher – klasa, która „owrapowuje” AASequence, zwraca najdłuższy znaleziony podciąg sekwencji wejściowej zgodny ze wzorcem

**3. Przykładowe dane wejściowe/wyjściowe**

Wzorzec zapisany w notacji PROSITE:

[RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYM]

Sekwencje testowe oraz ich podzbiory zgodne z powyższym wzorcem:

SRSLKMRGQAFVIFKEVSSAT

**RGQAFVIF**

KLTGRPRGVAFVRYNKREEAQ

**RGVAFVRY**

VGCSVHKGFAFVQYVNERNAR

**KGFAFVQY**