**Elementy Bioinformatyki**

**Projekt - sprawozdanie**

Zestawienia optymalne 2 sekwencji II

Anna Cholewczyńska 143198

Kamil Gersten 143220

Dominik Stefaniak 143133

# Opis programu

Po uruchomieniu programu, z pliku *input.txt* znajdującego się w tym samym folderze co program, pobrane zostają dwa wyrazy (sekwencje nukleotydowe).

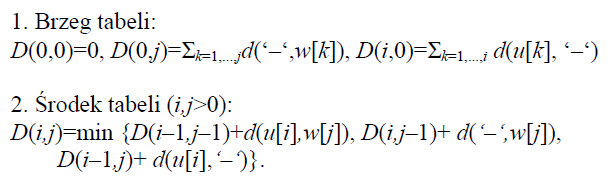
Następnie program oblicza tablicę odległości edycyjnych, oraz tablice A, B, C oraz S, potrzebne do wyznaczenia podobieństwa wyrazów.

Parametry takie jak wartość metryki, wartość podobieństwa znaków czy funkcja kary za przerwy są zdefiniowane w kodzie programu.

Następnie program wypisuje na konsolę wartość odległości edycyjnej sekwencji, ich dopasowanie, wartość podobieństwa wyrazów i ich optymalne dopasowanie.

# Zastosowane algorytmy

Przy wypełnianiu tablicy odległości edycyjnych D skorzystano z następujących wzorów:

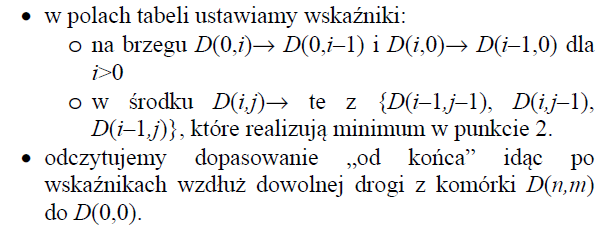


, gdzie *i* i *j* to długości słów odpowiednio pierwszego i drugiego.

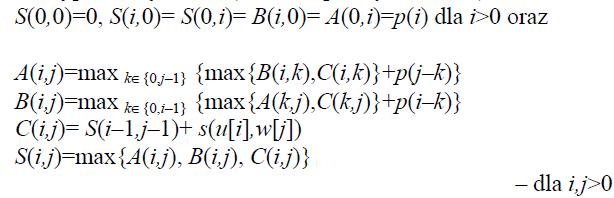
Funkcja *d(a,b)* zdefiniowana jest w ten sposób, że wynosi 0, gdy *a=b*, oraz 1 w innych przypadkach. Wartość tą można zmienić w kodzie programu.

Wartość odległości edycyjnej odczytywana jest z komórki D(i,j).

Optymalne dopasowanie wyznaczono za pomocą algorytmu



Dla wypełniania tablic A, B, C, S, skorzystano z wzorów:



, gdzie *p(x)* to funkcja kary za przerwy, zdefiniowana w kodzie.

Wartość podobieństwa sekwencji odczytywana jest z pola *S(i,j)*.

Dopasowanie optymalne dla podobieństw wyrazów wyznaczane jest poprzez wyznaczanie najdłuższej ścieżki wiodącej z *S(i,j)*, do *S(0,0)I.*