

Metody bioinformatyki - dokumentacja wstępna projektu

Michał Toporowski Marcin Swend

26 stycznia 2015

1 Cel projektu

Stworzenie interaktywnej aplikacji pozwalającej na zrozumienie działania algorytmu obliczającego macierz podobieństwa metodą BLOSUM.

2 Założenia

- aplikacja osadzona w przeglądarce
- wykorzystanie języka JavaScript i HTML5
- **dane wejściowe:** sekwencje wprowadzane przez użytkownika z klawiatury lub wczytywane z pliku
- **dane wyjściowe:** macierz podobieństwa dla danych sekwencji obliczona metodą BLOSUM oraz pośrednie wyniki obliczeń wyświetlane użytkownikowi poprzez przeglądarkę
- możliwość śledzenia przez użytkownika kolejnych kroków algorytmu obliczającego macierz
 - opis aktualnie wykonywanego kroku wraz z uzasadnieniem wyniku
 - wyróżnienie aktualnie obliczanej komórki/kolumny
 - możliwość przechodzenia do wybranych kroków i etapów algorytmu (bezpośrednie przejście do kolejnego etapu, powrót do poprzedniego kroku/etapu)

3 Algorytm

Wykorzystany zostanie algorytm składający się z następujących etapów:

1. **Obliczenie macierzy podstawień**

Obliczenie macierzy $A_{i,j}$ zawierającej częstości występowania symbolu i w danym miejscu na łańcuchu oraz symbolu j w tym samym miejscu w innych łańcuchach.

2. **Estymacja prawdopodobieństwa wystąpienia pary**

Obliczenie macierzy zawierającej prawdopodobieństwa wystąpienia pary $q_{i,j}$ na podstawie wzoru:

$$q_{i,j} = \frac{A_{i,j}}{\sum_{x,y} A_{x,y}}$$

3. **Estymacja prawdopodobieństw symboli**

Obliczenie prawdopodobieństw wystąpienia każdego z symboli p_i na podstawie wzoru:

$$p_i = q_{i,i} + \sum_{j \neq i} \frac{q_{i,j}}{2}$$

4. **Obliczenie macierzy BLOSUM**

Obliczenie macierzy $e_{i,j}$ na podstawie wzoru:

$$e_{i,j} = 2 \log_2 \left(\frac{q_{i,j}}{p_i p_j} \right)$$

Następnie obliczenie macierzy BLOSUM poprzez zaokrąglenie do liczb całkowitych.