Sprawozdanie 1

Ivan Bilunyk

Anastasiia Trubchaninova

**Temat:** Sekwencjonowanie łańcuchów DNA

Stosowane heurystyki są oparte na tabelach pokrycia słów. Przed rozpoczęciem poszukiwania rozwiązań wszystkie zdarzenia są porównywane ze sobą poprzez dopasowanie liter. Tabela zawiera informacje o pokryciu w przód (w prawo) i w tył (w lewo). Na przykład:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T | G | A | G | G | A | A | A | T | T |  |  |
|  |  | A | G | G | A | T | A | G | T | A | A |

Po otrzymaniu tabeli pokrycia algorytm „przegląda” kolejne słowa, dodając do wyniku nowo otrzymane litery. Kolejność słów jest wybierana na podstawie maksymalnego dostępnego zasięgu. Dzięki zastosowaniu dopasowania do przodu i wstecz, wyszukiwanie następnego słowa odbywa się „we wzorze szachownicy” - algorytm najpierw wybiera słowa o maksymalnym pokryciu z prawej strony, a gdy się kończą, przesuwa się na lewą stronę.

1) TGAGG**AAATT**   
2) **AAATT**GAAGC  
3) G**AAATT**GAAG

W tym przypadku :

* Pokrycie pary słów 3-2 wynosi 9 (w przód), analogicznie pary 2-3 wynosi 9 (wstecz).
* Para 3-1 ma pokrycie 6 (wstecz), a para 1-2 5 (w przód).

Niezbędne jest również zapobieganie cyklom - służy do tego szereg zastosowań - słowo użyte raz nie zostanie użyte ponownie w tym samym rozwiązaniu.

To rozwiązanie działa dokładnie dla idealnego zestawu. Jest również bardzo skuteczny w przypadku przypadkowych błędów pozytywnych i przypadkowych błędów negatywnych.

Przy braku pokrycia próbujemy ponownie dopasować z wykorzystaniem jednego z wcześniej użytych słów.