Marta Mazur 10.01.2018

Jakub Trzebiatowski

**Bioinformatyka**

**Sprawozdanie z projektu**

Zad 5. Zestawienia wielu sekwencji i profile. Typ B. Podobnie jak w przedostatnim: zademonstrować we własnym programie funkcje biblioteki służące do tworzenia i obróbki tym razem wielodopasowań i profili sekwencji.

Wybraną biblioteką do tworzenia i obróbki wielodopasowań i profili sekwencji jest BioJava. Wykorzystana została wersja 4.2.9. Wykorzystana została internetowa baza danych informacji o proteinach UniProt, dostępna pod adresem <http://www.uniprot.org/>.

Do pobierania sekwencji proteinowych z UnitProt stworzona została metoda:

*private static ProteinSequence getSequenceForId(String uniProtId)*

Metoda przyjmuje identyfikator sekwencji proteinowej, a zwraca pobrana już sekwencję. Sekwencje pobierane są w formacie Fasta z metody API dostępnej pod adresem [http://www.uniprot.org/uniprot/%s.fasta](http://www.uniprot.org/uniprot/%25s.fasta). Wykorzystująca klasę *FastaReaderHelper* umożliwiającą wczytywanie sekwencji proteinowych ze strumeni bądź plików.

Aby utworzyć profil sekwencji należy wywołać metodę *getMultipleSequenceAlignment* klasy *Alignments* podając jako parametr listę sekwencji:

*Alignments.getMultipleSequenceAlignment(lst);*

Metoda ta zwraca obiekt klasy *SimpleProfilePair* implementującej interfejs *Profile*.Interfejs Profile nie umożliwia późniejszej edycji wielodopasowania. Istnieje subinterfejs MutableProfile, natomiast w obecnej wersji biblioteki nie ma klasy go implementującej.

Dodatkowo można wyspecyfikować ustawienia takie jak

* *GapPenalty –* domyślną implementacją w BioJavie jest klasa SimpleGapPenalty. W której konstruktorze specyfikuje się open gap penalty oraz extension gap penalty. Obie wartości muszą być nieujemne:

*GapPenalty gapPenalty = new SimpleGapPenalty(1 , 3);*

* PairwiseSequenceScorerType – definiuje sposób przyznawania punków sekwencjom. Dozwolone wartości to:

*GLOBAL, GLOBAL\_IDENTITIES, GLOBAL\_SIMILARITIES, KMERS, LOCAL, LOCAL\_IDENTITIES, LOCAL\_SIMILARITIES, WU\_MANBER*

* ProfileProfileAlignerType – definiuje sposób dopasowania profile-profile. Dostępne wartości to:

*GLOBAL, GLOBAL\_CONSENSUS, GLOBAL\_LINEAR\_SPACE, LOCAL, LOCAL\_CONSENSUS, LOCAL\_LINEAR\_SPACE*

Jedyna sprawdzona przez nas możliwosć edycji utworzonego wcześniej profilu to skopiowanie z niego listy dopasowanych sekwencji (*profile.getAlignedSequences()*) i zmienienie dowolnej sekwencji w tej kopii na nową utworzoną „manualnie” (na podstawie poprzedniej dopasowanej sekwencji i listy kroków dopasowania) przez wywołanie:

*new* *SimpleAlignedSequence<ProteinSequence, AminoAcidCompound>(prev, steps);*

gdzie prev to poprzednio wygenerowana sekwencja a steps to lista kroków. Dostępne rodzaje kroków to

*AlignedSequence.Step.GAP, AlignedSequence.Step.COMPOUND*.