

Bioinformatyka 2019

Prowadzący dr Jacek Śmietański

Tematy projektów zaliczeniowych

Projekty mogą być realizowane w oparciu o jeden z poniższych scenariuszy:

1. Własny projekt – osoby zainteresowane proszę o przedstawienie propozycji celem akceptacji i ustalenia zakresu projektu.
2. Wsparcie projektu *biopython* (<https://github.com/biopython/biopython/issues>) lub implementacja użytecznej funkcjonalności, której w biopythonie brak – tu również osoby zainteresowane proszę o kontakt ze wskazaniem funkcjonalności, nad którą chcą pracować.
3. Rozwiązanie wybranych problemów algorytmicznych z platformy Rosalind. Szczegóły poniżej.

Ad. 3. Rosalind

Jeżeli chcesz rozwiązać problemy algorytmiczne z platformy Rosalind, wybierz 5 (pięć) spośród 22 wskazanych niżej problemów (dostępne również po zalogowaniu się do klasy na platformie):

- [Quartet Distance](#)
- [Wright-Fisher's Expected Behavior](#)
- [The Founder Effect and Genetic Drift](#)
- [Fixing an Inconsistent Character Set](#)
- [Assessing Assembly Quality with N50 and N75](#)
- [Genome Assembly with Perfect Coverage and Repeats](#)
- [Identifying Reversing Substitutions](#)
- [Isolating Symbols in Alignments](#)
- [Finding All Similar Motifs](#)
- [Overlap Alignment](#)
- [Semiglobal Alignment](#)
- [Enumerating Unrooted Binary Trees](#)
- [Counting Rooted Binary Trees](#)
- [Sex-Linked Inheritance](#)
- [Creating a Restriction Map](#)
- [Identifying Maximal Repeats](#)
- [Linguistic Complexity of a Genome](#)
- [Maximizing the Gap Symbols of an Optimal Alignment](#)
- [Using the Spectrum Graph to Infer Peptides](#)
- [Newick Format with Edge Weights](#)
- [Finding Disjoint Motifs in a Gene](#)

- [Wobble Bonding and RNA Secondary Structures](#)

Aby zapewnić różnorodność realizowanych zadań, obowiązują limity - pojedynczy problem może być realizowany przez maksymalnie 4-5 studentów.

Zasady wyboru projektu (Rosalind):

Należy przysłać mi (jacek.smietanski@ii.uj.edu.pl) **maila zawierającego listę zadań** (w postaci: nazwa zadania i jego 3-4 literowy kod), które chce się realizować, w kolejności od najbardziej do najmniej pożądanego.

Ze względu na ograniczenia i złożoną procedurę wyboru, zalecam by przesłać więcej niż 5 propozycji. Możesz też w mailu poprosić o losowy przydział zadań.

Tematy będą przydzielane z uwzględnieniem limitu (maksymalnie 4-5 osób realizujących jedno zadanie) oraz konieczności zagwarantowania realnego wyboru ostatniej zgłaszającej się osobie. Chciałbym również, aby każdy student realizował inny zestaw zadań. W przypadku, gdy wybrane zadanie nie może być przydzielone, przydzielam kolejne z listy aż do wybrania 5 zadań lub skończenia się listy z wyborem.

O przydzielonych zadaniach informuję mailowo niezwłocznie po dokonaniu przydziału. Jeżeli nie uda mi się przydzielić 5 zadań (wyczerpane limity), poproszę o uzupełnienie wyboru.

Zgłoszenia tematów można wysyłać od chwili publikacji tego ogłoszenia, nie należy jednak się spieszyć (proszę spokojnie zapoznać się z tematami zadań), gdyż pierwsze przydziały będą realizowane najwcześniej w **poniedziałek, 18.11.2019 o godzinie 8.00**. Zgłoszenia wysłane przed tym terminem będą rozpatrywane jako pierwsze, ale w obrębie tej grupy w losowej kolejności.

Zgłoszenia przesłane później (18.11 po godz. 8.00 lub w kolejnych dniach) będą rozpatrywane w kolejności nadsyłania.

Ostatecznym terminem wyboru tematów jest wtorek, 26.11.2019. Osobom, które do tego czasu nie wysłały zgłoszenia ze swoim wyborem lub wybór będzie zbyt mały, zadania zostaną przydzielone losowo.

Zasady realizacji zadań (Rosalind):

1. **Nie tworzymy repozytorium** i nie udostępniamy publicznie rozwiązań.
2. Nad rozwiązaniami pracujemy indywidualnie.
3. W sieci można znaleźć gotowe implementacje dla niektórych problemów. Proszę jednak się nimi nie inspirować i rozwiązywać zadania samodzielnie.
4. Kod zawierający rozwiązanie powinien być czytelny i dobrze udokumentowany.
5. Testy: do każdego zadania proszę o stworzenie własnego zestawu testów.
6. Wymagania podane w treści zadania proszę traktować jako minimum. Wskazane jest by rozwiązania były bardziej uniwersalne, np. akceptowały większe dane wejściowe, umożliwiały parametryzację algorytmów, zwracały wynik w rozszerzonej i bardziej czytelnej postaci.

7. Gotowy projekt należy mi przesłać w postaci paczki (.zip) zawierającej wszystkie skrypty oraz ew. pliki z danymi testowymi bądź rozwiązaniami, najpóźniej na dwa dni przed planowanym terminem obrony. Niezależnie od tego proszę realizowane problemy „zaliczyć” na platformie.