

Wstęp do programowania (studia stacjonarne)

Dr Anna Muranova

Semestr zimowy 2025/2026, UWM w Olsztynie

Ćwiczenie 12 (Pliki)

Zadanie 12.1. Wyobraź sobie, że jesteś bioinformatykiem i otrzymujesz kod genetyczny do analizy w pliku tekstowym.

Kod DNA składa się z 4 zasad azotowych: adeniny(A), cytozyny(D), guaniny(G), tyminy(T). Idealny kod DNA wygląda następująco:

TGCACGATCATGTCTACTATCCTCTCTATGGTGGGGTT...

Zdarza się, jednak, że kod zawiera małe jak i duże litery. Kolejny problem to maszyny sekwencjonujące nie są wolne od błędów. W zależności od maszyny błędy sekwencjonowania mogą zostać zamienione na znak – czy literę N. W dokumentacji znajduje się następujący zapis:

gdy jakoś sekwencji nie pozwala dokładnie odczytać rodzaju zasady azotowej wstawiany jest znak „-” Natomiast, gdy laser czytujący ześlizgnie się, wstawiane są litery „N”, jednocześnie ostatnia wartość zasady jest ponownie odczytywana bez ubytku zasady w tym miejscu.

- (a) wczytaj plik <http://wmii.uwm.edu.pl/~muranova/WdPSt2025-26/DNA.txt>
- (b) policz ile razy występuje w kodzie każda zasada azotowa - adenina, cytozyna, guanina, tymina.
- (c) Oczyść DNA z błędów typu N.
- (d) Policz wystąpienia sekwencji GAGA
- (e) Znajdź miejsce (indeks) w łańcuchu, gdzie występuje 7 guanin z rzędu
- (f) Znajdź miejsce (indeks) , gdzie od końca łańcucha występuje 6 cytozyn
- (g) Policz ile razy w kodzie pojawiła się sekwencja CTGAAA
- (h) W sekwencji CTGAAA czasem mutuje ostanía litera A, wówczas jakoś ostatniej litery może być wątplia. Ile sekwencji znajdziesz, jeśli weźmiesz pod uwagę wątpliwą, ostatnią adeninę?
- (i) Na podstawie czystej nici utwórz odpowiadającą jej nić RNA (nić RNA w miejscu tyminy będzie mieć uracyl (U)). Nic RNA zapisz do nowego pliku RNA.txt

Zadanie 12.2. Można używać bibliotekę `random`

- (a) Napisz program, który w pętli generuje 100 liczb całkowitych od -100 do 100 i zapisuje każdą z tych liczbę do odpowiedniego pliku: parzyste do even.txt, a nieparzyste do odd.txt.
- (b) Napisz program, który wczyta pliki utworzone w poprzednim zadaniu i obliczy sumę oraz średnią arytmetyczną liczb znajdujących się w każdym z nich.