

## Sortowanie kodonów genomu

### Polecenie:

Napisz algorytm sortowania listy kodonów stworzonej z fragmentów DNA pochodzących z jego sekwencjonowania w oparciu o iteracyjną wersję algorytmu *quicksort*.

### Wejście:

Pierwszą podawaną wartością będzie dodatnia liczba całkowita oznaczająca ilość zestawów danych, po której na wejściu pojawią się zestawy danych w ilości równej wczytanej liczbie.

Pojedynczy test składa się z następujących pól:

- sekwencji znaków (*String*) *s*, która jest wynikiem sekwencjonowania genomu.

### Wyjście:

Posortowany genom bez sekwencji startowej i końcowej

### Założenia i wymagania:

1. W ciągu znaków z sekwencjonowania nie powinno być innych liter niż "A", "C", "G", "T" oznaczających aminokwasy (*adenine*, *cytosine*, *guanine*, *thymine*). Jeżeli występują inne znaki to program powinien wypisać na ekran: *"Wrong character in DNA sequence."*
2. Kodon to ciąg 3 aminokwasów.
3. Początek genomu jest oznaczany przez kodon "ATG", natomiast jego koniec przez jedną z trzech sekwencji: "TAA", "TGA" lub "TAG". Sekwencja początkowa występuje przed sekwencją końcową czytając od lewej do prawej. Natomiast przed sekwencją startową jak i po tej sekwencji możliwe jest wystąpienie pojedynczych aminokwasów, które odrzuca się w dalszej analizie. Np.: AC**ATG**AAAT**GAC** - pogrubiona część oznacza poprawną sekwencję do dalszej analizy.
4. Pomiedzy sekwencją startową a sekwencją końcową nie powinny wystąpić ponownie sekwencje końcowe bądź startowa. Jeśli tak się stanie program powinien zwrócić: *"More than one START/STOP codon."*

5. Jeżeli program nie znajdzie sekwencji startowej/końcowej powinien wypisać "Wrong DNA sequence."
6. Pomiedzy sekwencją startową a sekwencją końcową liczba aminokwasów powinna być wielokrotnością trójki, by móc poprawnie rozpoznać kodony. W przeciwnym wypadku program powinien wypisać na ekran: "Wrong DNA sequence."
7. Kodony należy posortować alfabetycznie (od A do T).
8. Program na zakończenie powinien wypisywać posortowany genom ze względu na kodony.
9. Do sortowania kodonów genomu należy wykorzystać iteracyjny algorytm *quickSort*, w którym **nie zostaną wykorzystane stosy oraz rekurencja**.
10. Dozwolone jest wykorzystanie do **4 tablic na poziom sortowania** jako zmiennych pomocniczych, tzn.: Wpierw sortowanie odbywa się po pierwszych literach kodonów, a następnie po drugich literach kodonów i na końcu po trzecich literach kodonów.
11. Każdy z poziomów sortowania powinien korzystać z tej samej funkcji sortującej. **Funkcja sortująca powinna być wyodrębniona w kodzie**.
12. Niedozwolone jest używanie innego kodowania znaków niż ASCII.
13. Program powinien być niewrażliwy na wielkość liter.
14. Jedyny dozwolony import to `java.util.Scanner`.
15. Algorytm sortowania nie powinien się wykonywać jeśli choć jeden z warunków dotyczących sekwencji jest niespełniony.

#### Tło zadania:

Podstawowym elementem analizy genomu jest jego analiza bioinformatyczna polegająca na rozpoznaniu i przetworzeniu informacji o kodonach, ich liczności, występowaniu, statystykach w taki sposób by później można było dokonać analizy pomiędzy genomem referencyjnym a genomem analizowanym. Jeżeli następują różnice na poziomie genomu referencyjnego i genomu analizowanego to można mieć wtedy do czynienia z mutacją. Wykrywanie tych mutacji jest bardzo ważne ze względu na późniejsze identyfikowanie choroby i doboru leczenia. Np.: w przypadku raka piersi taką mutacją jest mutacja genów BRCA1 i BRCA2, która znacząco zwiększa prawdopodobieństwo choroby. Najbardziej znanym przypadkiem profilaktycznym jest przypadek Angeliny Jolie, u której takie mutacje zostały zdiagnozowane, a sama kobieta poddała się odpowiednim zabiegom zapobiegawczym.

Jako, że analizy genomu są znacznie bardziej skomplikowane niż wyszukanie średniego kodonu, to dla wielu z nich jest jeden krok wspólny, a mianowicie sortowanie, dla którego to ćwiczenie to jest dedykowane.

**Przykładowe testy:**

Wejście	Wyjście
ATGAACAAATGA	AAAAAC
ATGAAAATGAACTGA	More than one START/STOP codon.
AACTTTAAATGA	Wrong DNA sequence.
ATGWWWTGA	Wrong character in DNA sequence.