

Algorytmy i Struktury Danych 2: Laboratorium 14

Maciej Pióro

Kocia choroba

Wkrótce po swoim powrocie z dalekich podróży Fryderyk zauważył, że jego koty zaczęły utykać na lewą przednią łapkę. Chcąc sprawdzić podejrzenie, że ta choroba jest spowodowana przywleczonym przez niego z zagranicy patogenem, Fryderyk zsekwencjonował ciągi aminokwasów białek znajdujących się na sierści jego milusińskich. Niestety po zapisaniu wyników eksperymentu jeden z kotków, Bonifacy, zrzucił na jego komputer magnes, przez co niektóre fragmenty sekwencji na dysku twardym zreplikowały się w innych miejscach. Fryderyk chce naprawić dane poprzez usunięcie zduplikowanych fragmentów. Twoim zadaniem będzie znalezienie najdłuższego fragmentu sekwencji (spójnego podciągu), który pojawia się przynajmniej dwukrotnie.

Zadanie pierwsze

Na wejściu do programu otrzymujemy pojedynczy łańcuch znaków: `text`. Naszym zadaniem jest znalezienie najdłuższego fragmentu sekwencji, który powtarza się przynajmniej dwukrotnie.

Uwagi

- Wymagana złożoność czasowa: $\mathcal{O}(n^2)$, gdzie n to długość wejściowej sekwencji
- Oba wystąpienia fragmentu muszą być rozłączne, jakkolwiek nie muszą występować bezpośrednio po sobie (choć mogą)
- W przypadku kiedy nie ma żadnego powtarzającego się fragmentu, za najdłuższy taki fragment uznajemy pusty string
- W rozwiązaniu jest sugerowane użycie techniki programowania dynamicznego lub zainspirowanie się algorytmem KMP
- W przypadku istnienia wielu (równie długich) rozwiązań można zwrócić dowolne z nich

Szczepionka na kocią chorobę

Gdy Fryderykowi udało się już naprawić sekwencje, przystąpił do opracowania szczepionki. W tym celu postanowił sprawdzić, z jakimi znanymi mu patogenami może być spokrewniony wirus wywołujący kocią chorobę. Do tego będzie mu potrzebny program, który przyjmuje dwie sekwencje białkowe i identyfikuje najdłuższy wspólny fragment (spójny podciąg) obu tych sekwencji.

Zadanie drugie

Na wejściu do programu otrzymujemy dwa łańcuchy znaków: `x` oraz `y`. Naszym zadaniem jest znalezienie najdłuższego fragmentu sekwencji, który powtarza się w obu łańcuchach.

Uwagi

- Wymagana złożoność czasowa: $\mathcal{O}(mn)$, gdzie m, n to długości sekwencji `x, y`
- W przypadku kiedy nie ma żadnego wspólnego fragmentu, za najdłuższy taki fragment uznajemy pusty string
- Rozwiązanie może być zupełnie niezależne od rozwiązania zadania pierwszego
- W przypadku istnienia wielu (równie długich) rozwiązań można zwrócić dowolne z nich

Etapy

Etap 1 - pierwsze zadanie, zwrócić długość najdłuższego powtarzającego się fragmentu (0.5 pkt)

Etap 2 - pierwsze zadanie, zwrócić najdłuższy powtarzający się fragment (0.5 pkt)

Etap 3 - drugie zadanie, zwrócić długość najdłuższego fragmentu występującego w obu sekwencjach (1 pkt)

Etap 4 - drugie zadanie, zwrócić najdłuższy fragment występujący w obu sekwencjach (0.5 pkt)