

Odległość Levenshteina

Struktura

Client.java

Metoda main()

Główna metoda programu, odczytuje parametry wejściowe którymi są: lokalizacja pliku tekstowego z danymi oraz ciąg znaków dla którego będzie przeprowadzone wyszukiwanie najbardziej zbliżonego ciągu. Funkcja pobiera z konsoli dane za pomocą `String[] args`, Dlatego muszą one być podane w dobrej kolejności. Następnie dane z pliku tekstowego są wczytywane do listy za pomocą klasy `java.nio.file.Files`. Po wczytaniu dla wszystkich elementów listy wyliczana jest odległość Levenshteina, podczas iterowania listy zapamiętywana jest najmniejsza wartość odległości oraz numer indeksu w liście elementu z najmniejszą odległością. Po przeszukaniu całej listy wypisywany jest indeks elementu z najmniejszą miarą Levenshteina. Aby indeks był równy numerowi linii w którym znajduje się dany ciąg znaków w pliku tekstowym, do indeksu jest dodawana jedynka.

Levenshtein.java

Metoda minimalDistance()

Metoda statyczna która zwraca wartości typu `int`. Jako parametry wejściowe przyjmuje dwie wartości typu `String` odpowiadające dwóm porównywanym ciągami znaków. Na początku algorytmu tworzona jest tablica dwuwymiarowa typu `int` o rozmiarze $n+1 \times m+1$ gdzie n i m są długościami porównywanych ze sobą ciągów znaków. Pierwszy wiersz i pierwsza kolumna są wypełniane wartościami od 0 do $n+1$ oraz 0 do $m+1$, ponieważ odpowiada to ilości wykonanych wstawień i usunięć z ciągu. Następnie w pętli dwupoziomowej dokonywane jest porównanie znaku dla danego wiersza i danej kolumny jeżeli znaki są takie same zwraca 0 w przeciwnym razie jej wartość wynosi 1. Aby ustalić wartość odległości w danej komórce wyznaczane jest minimum z trzech liczb:

- wartości komórki leżącej bezpośrednio nad aktualną komórką zwiększonej o 1,

- wartości komórki leżącej bezpośrednio w lewo od aktualnej komórki + 1
- wartości komórki leżącej bezpośrednio w lewą-górną stronę od aktualnej komórki + *cost*

Po przejściu przez pętle i wypełnieniu całej tablicy odległość edycyjna znajduje się w komórce ostatniego wiersza i ostatniej kolumny. Ta liczba jest zwracana z funkcji jako wynik porównania.