

Algorytmy i struktury danych, Teleinformatyka, I rok

Raport z laboratorium nr: 5

Imię i nazwisko studenta: Filip Komarzyniec

nr indeksu: 296913

1. W pole poniżej wklej najważniejszy (według Ciebie) fragment kodu źródłowego z zajęć (maksymalnie 15 linii).

```
1. with open(„1000_pattern.txt”, „r”) as f: #efektywne wczytywanie pliku w pythonie
2.     f = f.readlines()
3.     FILE = [letter.strip(„\n”) for letter in f]
4.
5. def Naive_Matcher(FILE, TEMPL): #definicja funkcji naiwnego wyszukiwania
6.     s, count = 0, 0
7.     n = CcharF(FILE) #Cchar() to funkcja licząca długość wiersza w FILE, czyli podanym pliku
8.     m = len(TEMPL) #TEMPL to wzorzec
9.     C = sum(1 for k in FILE) #liczę ilość wierszy w FILE
10. for i in range(s, n-m+1): #iteruję po kolumnach
11.     for j in range(0, C-2): #iteruję po wierszach
12.         if TEMPL[0][0:m] == FILE[j][i:i+m] and TEMPL[1][0] == FILE[j+1][i] and
13.             TEMPL[2][0] == FILE[j+2][i] #warunki znalezienia wzorca w FILE
```

Uzasadnij swój wybór.

Powyżej wstawione przeze mnie fragmenty kodu uważam za istotne, ponieważ rozwiązują one zasadnicze problemy zadanego ćwiczenia, tj. efektywne wczytanie otrzymanego pliku tekstowego do dwuwymiarowej tablicy oraz wyszukiwanie w niej ustalonego wcześniej, dwuwymiarowego wzorca.

2. Podsumuj wyniki uzyskane podczas wykonywania ćwiczenia. Jeśli instrukcja zawierała pytania, odpowiedz na nie.

Zadany wzorzec w macierzach o rozmiarach (N) 1000x1000, 4000x4000 oraz 5000x5000 znalazł się odpowiednio 6, 97 i 161 razy. Zmierzone przeze mnie czasy jego wyszukiwania dla tych macierzy, przy użyciu algorytmów naiwnego i Karpa-Rabina, prezentują się następująco :

	<u>Alg naiwny</u>		<u>Alg Karpa-Rabina</u>	
czas dla N = 1000	0,45	[s]	1,06	[s]
czas dla N = 4000	9,85	[s]	18,96	[s]
czas dla N = 5000	15,58	[s]	29,38	[s]

Algorytm hashujący Karpa-Rabina okazał się około dwa razy wolniejszy w działaniu od algorytmu naiwnego. Wpływ na taki stan rzeczy ma przede wszystkim złożoność obliczeniowa obu algorytmów, która w pesymistycznym przypadku wynosi dla pierwszego z nich $O(m \cdot n)$, dla drugiego $O(m \cdot (m + n + 1))$ oraz fakt, iż krótkiego wzorca wyszukiwałem w dwóch wymiarach. Dla długiego, jednowymiarowego ciągu znaków algorytm hashujący mógłby być szybszy.

Raporty należy wysłać na adres: andmat+aisd@agh.edu.pl . Termin oddania raportów to 7 dni po zajęciach laboratoryjnych, których dotyczy raport.