

Autor: Hubert Kowalczyk 259550	Struktury Danych i złożoność obliczeniowa Semestr letni 2022/2023	Termin: Wtorek NP: 17:05
Prowadzący: Dr. Inż. Tomasz Kapłon	Ćwiczenie 4	Data wykonania ćwiczenia: 06.06.2023
		Data oddania sprawozdania 20.06.2023

1) Cel Ćwiczenia

Celem ćwiczenia było zaimplementowanie algorytmu służącego do wyszukiwania wzorców w tekście. Pośród zaproponowanych przez prowadzącego algorytmów wybrano algorytm Knutha-Morrisa-Pratta

2) Algorytm i implementacja

2.1) Algorytm

Algorytm Knutha-Morrisa-Pratta jest algorytmem wyszukiwania wzorca w tekście. Polega na przesuwaniu wzorca wzdłuż tekstu i pomijaniu niepotrzebnych porównań znaków, wykorzystując tablice LPS. Tablica LPS zawiera informacje o najdłuższym prefiksie wzorca.

2.1) Implementacja

Funkcja 'computeLPS' jest odpowiedzialna za obliczanie tablicy LPS dla wzorca. Tablica LPS przechowuje informacje o najdłuższym prefiksie wzorca, który jest jednocześnie sufiksem

Funkcja 'KMP' implementuje sam algorytm Knutha-Morrisa-Pratta, który wyszukuje wzorec w tekście. Inicjalizuje ona tablice LPS dla wzorca i tworzy zmienne pomocnicze 'i', 'j', 'lineNum' i 'found'. Indeksy 'i' i 'j' używane są do przechodzenia przez tekst i wzorec. Zmienna 'lineNum' śledzi aktualny numer linii, a 'found' informuje, czy wzorec został odnaleziony w tekście.

2.3) Plik config.cfg

Plik config posiada pola do wybrania pliku, w którym znajduje się badany tekst (Plik z danymi). Pole Znalezione wzorce to plik.csv, w którym znajdują się informacje, w których liniach znaleziono wzorec

lub też informacja o nieznalezieniu wzorca. Standardowo znajdują się też nazwa algorytmu. Ponadto w pliku config można wpisać wzorzec do wyszukiwania oraz ścieżkę pliku z zmierzonymi czasami wykonywania algorytmu.

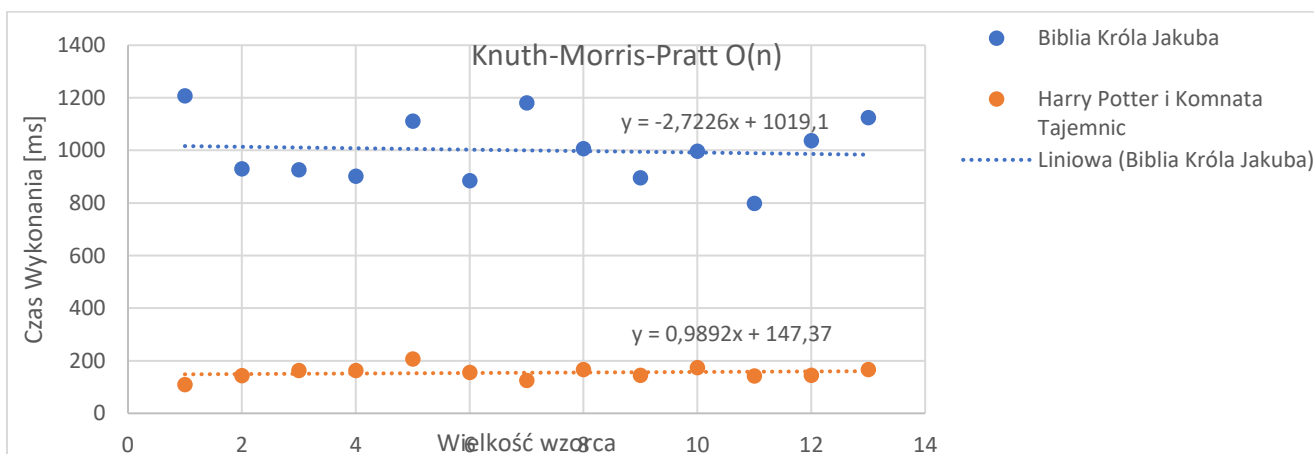
2.4) Metodologia badawcza

Algorytm badano na dwóch książkach w języku angielskim. Pierwsza z nich to biblia króla Jakuba w postaci pliku .txt (4,5mb około 30000linii tekstu) oraz książkę Harry Potter i komnata Tajemnic w formacie .txt(500kb tekstu oraz około 14000 linii tekstu).Dla tych dwóch tekstów badano jak zmieni się czas wykonywania w zależności od długości wzorca dla długości 1-13.Czas dla każdej długości badano 10 razy a następnie liczono średnią, która posłużyła jako wynik dla danej instancji. Algorytm badano na komputerze wyposażonej w procesor Intel Core i7-9750H.

3) Wyniki

Tabela 1 Tabela w której zgromadzono uzyskane wyniki pomiarów

Biblia Króla Jakuba		Harry Potter i Komnata Tajemnic	
Wielkość Instancji	Czas Wykonania [ms]	Wielkość Instancji	Czas Wykonania [ms]
1	1207,15	1	109,92
2	929,70	2	143,31
3	926,71	3	162,97
4	901,44	4	162,83
5	1111,10	5	207,18
6	884,98	6	155,81
7	1181,17	7	125,26
8	1006,26	8	167,00
9	895,46	9	144,29
10	996,54	10	174,28
11	798,69	11	142,60
12	1037,03	12	144,40
13	1124,59	13	165,97



Wykres 1 Wykres sporządzony na podstawie wyników z tabeli 1

