

Sprawozdanie z laboratorium 1 – wyszukiwanie wzorca w tekście

0. Zaimplementuj algorytmy wyszukiwania wzorców

Zostały zaimplementowane oraz sprawdzone pod względem poprawności działania następujące algorytmy: naiwny, automat skończony, algorytm KMP

1. Zaimplementuj testy porównujące szybkość działania wyżej wymienionych algorytmów.

Została zaimplementowana funkcja porównująca czas działania algorytmów dla określonych danych wejściowych.

2. Znajdź wszystkie wystąpienia wzorca "art" w załączonej ustawie, za pomocą każdego algorytmu.

Każdy z 3 algorytmów dał następujący wynik wyszukiwania:

```
Znaleziono 273 powtórzeń
[1156, 1505, 4692, 4734, 4879, 5082, 5148, 5949, 6039, 7266, 7511, 7781, 8044, 8299, 9104, 9959, 10022, 10224, 11122, 1120
7, 11618, 13194, 15284, 15358, 16092, 16261, 16406, 16547, 16616, 16840, 16856, 23637, 24061, 24152, 24586, 24683, 24780, 2
4931, 25530, 25689, 27001, 27288, 27479, 27542, 27592, 27857, 28373, 28558, 28766, 30964, 31021, 31096, 31362, 31811, 3260
9, 32968, 33053, 33268, 33595, 34651, 34737, 35511, 36155, 37143, 37543, 38451, 38595, 39056, 39210, 39436, 39568, 39980, 4
1152, 41829, 42028, 42198, 42371, 42504, 42718, 42896, 42941, 43447, 43555, 43787, 44590, 44653, 44953, 45010, 45293, 4540
1, 47319, 47422, 48785, 48820, 48906, 49052, 49259, 49316, 49488, 49559, 49915, 49979, 50102, 50160, 50702, 51050, 51179, 5
1966, 52071, 52272, 52552, 53008, 53032, 53211, 53788, 53931, 54078, 54137, 54770, 55075, 55279, 55465, 55807, 55991, 5682
7, 56911, 57164, 57549, 57800, 57932, 57989, 58280, 58378, 58874, 58966, 59395, 59523, 59949, 60296, 60549, 60794, 61262, 6
1770, 62463, 62610, 62663, 63504, 63698, 63789, 63869, 65075, 65114, 65176, 66024, 66330, 66982, 67056, 67578, 67695, 6772
0, 67852, 67926, 68226, 68854, 69046, 69172, 69249, 69604, 69697, 69794, 70109, 70512, 70664, 70698, 71561, 71705, 72106, 7
2314, 73115, 74253, 75346, 75470, 75486, 75493, 75717, 75803, 75871, 76247, 77030, 78047, 78558, 78872, 78936, 78970, 7909
1, 79270, 79341, 79410, 79448, 79582, 79822, 79884, 80848, 83234, 84578, 84828, 85834, 86793, 86819, 87093, 87174, 87359, 8
7399, 87640, 87707, 87947, 88033, 88073, 88136, 88391, 88665, 88940, 89047, 89316, 89325, 89341, 89361, 90167, 90218, 9057
2, 91682, 91795, 92916, 93366, 93426, 93561, 94603, 94808, 95981, 96120, 97285, 98770, 99825, 102953, 104141, 104723, 10576
7, 105973, 110186, 115010, 115163, 116109, 144052, 158607, 159478, 161581, 162788, 163969, 168897, 169074, 178453, 185631,
200530, 200628, 202696, 206794, 209181, 211819, 212319, 217444, 217891, 223168, 223252]
```

3. Porównaj szybkość działania algorytmów dla problemu z p. 3.

Dla danego problemu, czasy działania algorytmów automatu skończonego oraz KMP są do siebie porównywalne oraz znacznie krótsze od czasu algorytmu naiwnego.

```
{'Naive      ': 107015}
{'Automata   ': 90018}
{'KMP        ': 88177}
```

Czas wyrażony w mikrosekundach

4. *Zaproponuj tekst oraz wzorzec, dla którego zmierzony czas działania algorytmów (uwzględniający tylko dopasowanie, bez pre-processingu) automatu skończonego oraz KMP będzie co najmniej 5-krotnie krótszy niż dla algorytmu naiwnego.*

Dla następujących danych:

```
pattern = 'a'*100000
text = ('a'*99999 + 'b')*10
```

Czasy działania algorytmów są następujące

```
{'Naive': 8125308}
{'Automata': 419999}
{'KMP': 797390}
```

Czas wyrażony w mikrosekundach

Dane więc spełniają warunek. Jest to przypadek najmniej optymalny dla algorytmu naiwnego ponieważ litera różniąca wzorzec od tekstu znajduje się na końcu wzorca.

5. *Zaproponuj wzorzec, dla którego zmierzony czas obliczenia tablicy przejścia automatu skończonego będzie co najmniej 5-krotnie dłuższy, niż czas potrzebny na utworzenie funkcji przejścia w algorytmie KMP.*

Taki wzorzec to każde wyrażenie którego liczność zbioru alfabetu jest zbliżona do długości wzorca. Czasy działania funkcji dla wzorca `string.ascii_lowercase`, czyli zbiór małych liter alfabetu łacińskiego, przedstawiają się następująco:

```
{'transition_table': 8010}
{'prefix_function': 0}
```

Czas wyrażony w mikrosekundach