

Laboratorium 1

Weronika Ormaniec

Marzec 2020

1 Analiza fragmentu Wikipedii

Przeanalizowano pierwsze 10^8 B dostarczonego pliku, z uwagi na to, że przetworzenie całości linia po linii było zbyt czasochłonne. Poniżej czas pracy poszczególnych algorytmów i odpowiadające im dopasowania słowa "kruszwil".

- Naive algorithm: 20.68013310432434
- Automata algorithm (with transition table generation): 32.09707474708557
- KMP algorithm (with prefix function generation): 17.271249294281006

Znalezione dopasowania:

- Naive algorithm: 2 [42549966, 50733405]
- Automata algorithm: 2 [42549966, 50733405]
- KMP algorithm: 2 [42549966, 50733405]

2 Analiza ustawy

- Naive algorithm: 0.039752960205078125
- Automata algorithm (with transition table generation): 0.06156778335571289
- KMP algorithm (with prefix function generation): 0.039055585861206055

Poniżej otrzymane dopasowania (liczba i pierwsze 10 z nich):

- Naive algorithm: 273 [844, 1193, 4302, 4344, 4489, 4692, 4758, 5442, 5532, 6759]
- Automata algorithm: 273 [844, 1193, 4302, 4344, 4489, 4692, 4758, 5442, 5532, 6759]
- KMP algorithm: 273 [844, 1193, 4302, 4344, 4489, 4692, 4758, 5442, 5532, 6759]

3 Przypadki skrajne

Tekst oraz wzorzec, dla którego zmierzony czas działania algorytmów 2 oraz 3 będzie co najmniej 2-krotnie krótszy niż dla algorytmu naiwnego

Dla poniższego przypadku widać kwadratową naturę złożoności algorytmu naiwnego, który po zakończeniu porównania pierwszego wystąpienia wzorca, będzie wracał do początku wzorca, podczas gdy pozostałe algorytmy nie. W tekście wzorzec występuje przy każdym przesunięciu.

Tekst = a^{10^6} Wzorzec = a^{10^5}

- Naive algorithm: 10.248653411865234
- Automata algorithm: 0.4861330986022949
- KMP algorithm: 0.4329671859741211

Poniżej otrzymane dopasowania (liczba i pierwsze 10 z nich):

- Naive algorithm: 900001 [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
- Automata algorithm: 900001 [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
- KMP algorithm: 900001 [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

Wzorzec, dla którego zmierzony czas obliczenia tablicy przejścia automatu skończonego będzie co najmniej dwukrotnie dłuższy, niż czas potrzebny na utworzenie funkcji przejścia w algorytmie KMP

Wystarczyło wybrać dowolny długi wzorzec, gdyż przetwarzanie wzorca dla automatu zależy sześciennie od jego długości, a dla KMP liniowo od jego długości.

Wzorzec = $(abcdefghijklmnopqrstuvwxyz)^{10}$

- Automata transition table generation: 1.5897703170776367
- KMP prefix function generation: 5.602836608886719e-05