

Algorytmy Tekstowe

Laboratorium 6 - Wyszukiwanie wzorca dwuwymiarowego

26.05.2021

Magdalena Pastuła

Zadanie 2

Otrzymane następujące wyniki, gdzie indeksowanie zaczyna się od 1:

```
"n": [(33, 2), (3, 10), (58, 14), (37, 19), (66, 30), (53, 33), (56, 34), (69, 36), (21, 38), (69, 41), (16, 55), (22, 57), (69, 58), (23, 63), (2, 84)]
"x": [(30, 69)]
"o": [(43, 2), (55, 2), (52, 3), (54, 9), (81, 11), (35, 12), (29, 18), (30, 18), (35, 27), (12, 28), (34, 35), (8, 39), (9, 39), (73, 43), (60, 46), (83, 53), (46, 56), (32, 59), (17, 61), (7, 67), (6, 76)]
"t": [(39, 1), (52, 1), (18, 4), (73, 4), (74, 4), (25, 5), (26, 5), (71, 6), (3, 7), (2, 8), (3, 8), (24, 9), (37, 11), (74, 11), (56, 12), (17, 13), (6, 15), (32, 17), (79, 23), (6, 24), (30, 24), (48, 25), (9, 30), (29, 32), (21, 34), (53, 34), (61, 34), (5, 38), (43, 46), (60, 50), (30, 53), (57, 55), (15, 56), (63, 57), (74, 60), (54, 62), (69, 72), (43, 74), (10, 76), (61, 76), (60, 79)]
"h": [(29, 3), (39, 3), (75, 13), (58, 32)]
"r": [(3, 5), (54, 6), (35, 11), (9, 14), (19, 15), (17, 19), (71, 23), (45, 26), (69, 30), (62, 31), (35, 38), (49, 38), (8, 40), (64, 40), (57, 41), (48, 43), (8, 51), (21, 55), (22, 55), (30, 66), (33, 71)]
"i": [(33, 1), (3, 6), (75, 14), (79, 14), (57, 18), (33, 32), (46, 34), (10, 38), (62, 46), (70, 52), (21, 56), (11, 61), (54, 70)]
"s": [(56, 1), (51, 15), (10, 22), (73, 25), (81, 25), (39, 35), (47, 35), (69, 38), (72, 42), (48, 45), (30, 46), (6, 50), (54, 54), (31, 57), (32, 57), (5, 58), (11, 59), (5, 64), (42, 64)]
"d": [(39, 20)]
"l": [(35, 46), (55, 46), (48, 62), (30, 73), (43, 78)]
"e": [(12, 2), (16, 3), (26, 4), (19, 7), (78, 7), (79, 7), (82, 7), (3, 9), (22, 11), (42, 12), (83, 15), (71, 16), (69, 18), (74, 24), (42, 27), (20, 28), (75, 28), (53, 32), (44, 37), (31, 39), (73, 39), (17, 44), (31, 44), (70, 47), (84, 48), (39, 49), (44, 49), (72, 50), (49, 51), (60, 51), (48, 53), (24, 54), (59, 55), (60, 55), (43, 58), (23, 62), (2, 64), (12, 65), (9, 66), (26, 66), (80, 66), (65, 67), (30, 68), (67, 70), (68, 73), (30, 74), (61, 74), (6, 78)]
"w": [(3, 4), (23, 71)]
"f": [(79, 2), (32, 60)]
"y": [(46, 6)]
"c": [(43, 1), (70, 1), (15, 11), (84, 42), (12, 46), (5, 55)]
"a": [(66, 3), (39, 5), (22, 7), (58, 12), (54, 13), (55, 13), (66, 15), (78, 22), (66, 23), (61, 25), (5, 31), (67, 36), (71, 36), (59, 37), (60, 37), (81, 38), (79, 43), (55, 49), (33, 51), (80, 60), (7, 61), (79, 62), (8, 64), (35, 67), (30, 70), (33, 74), (78, 75), (2, 83)]
"p": [(43, 19), (30, 72)]
"m": [(46, 1), (18, 6), (36, 41), (36, 61), (30, 71)]
```

Liczba wszystkich takich znalezionych wzorców wyniosła 235. Aby każda linia miała tą samą długość, dodano na koniec znaki '0', które następnie zostały zignorowane.

Zadanie 3

Nie znaleziono żadnych wystąpień wzorca "th" w dwóch kolejnych liniach na tej samej pozycji.

Znaleziono tylko jedno wystąpienie wzorca "t h" w dwóch liniach na tej samej pozycji, a dokładnie: (39, 3).

Zadanie 4

Wybrano litery "k", "o" i "s".

Poniższa tabela przedstawia liczbę znalezionych wystąpień tych liter i faktyczną liczbę liter znalezionych w pliku tekstowym.

| Wzorzec | Liczba znalezionych | Faktyczna liczba |
|---------|---------------------|------------------|
| k | 12 | 22 |
| o | 308 | 369 |
| s | 277 | 334 |

Niezgodności najprawdopodobniej wynikają ze złego przycięcia wzorców liter przy wycinaniu z oryginalnego obrazu.

Zadanie 5

Znalezione wystąpienia wzorca "p a t t e r n":

```
"p a t t e r n": [(603, 243), (493, 285), (559, 343), (647, 349), (515, 529)]
```

Zadanie 6

Poniższa tabela przedstawia zmierzone czasy budowania automatów i czasy wyszukiwania wzorców dla poszczególnych wzorców:

| Wzorzec | Czas przygotowania | Czas przeszukiwania |
|---------------|--------------------|---------------------|
| th | 0.0662 ms | 4.7232 ms |
| t h | 0.0596 ms | 4.763 ms |
| k | 0.9299 ms | 1.193 s |
| o | 1.1206 ms | 1.223 s |
| s | 1.2917 ms | 1.219 s |
| p a t t e r n | 27.100 ms | 1.135 s |

Jak można zauważyć, wyszukiwanie wzorca w obrazie trwa mniej więcej tyle samo, niezależnie od wielkości wzorca. Natomiast dla wzorca "p a t t e r n" znacząco wzrósł czas przygotowania do wyszukiwania, czyli przygotowanie automatu do wyszukiwania wzorców w jednym wymiarze i automatu do wyszukiwania odpowiedniej kolejności tych wzorców.

Zadanie 7

Czasy przeszukiwania wzorców przy odpowiednich podziałach:

| Wzorzec | Liczba fragmentów | Czas przeszukiwania |
|---------------|-------------------|---------------------|
| th | 2 | 4.7359 ms |
| th | 4 | 4.8537 ms |
| th | 8 | 5.1282 ms |
| t h | 2 | 4,6213 ms |
| t h | 4 | 4,9989 ms |
| t h | 8 | 4,9705 ms |
| k | 2 | 1.155 s |
| k | 4 | 1.140 s |
| k | 8 | 1.139 s |
| o | 2 | 1.211 s |
| o | 4 | 1.187 s |
| o | 8 | 1.154 s |
| s | 2 | 1.157 s |
| s | 4 | 1.158 s |
| s | 8 | 1.163 s |
| p a t t e r n | 2 | 1.089 s |
| p a t t e r n | 4 | 1.075 s |
| p a t t e r n | 8 | 1.074 s |

Jak można zauważyć, im większy podział, tym więcej czasu zajmuje przeszukiwanie w przypadku przeszukiwania w pliku tekstowym. Natomiast w przypadku wyszukiwania wzorca w obrazie można zauważyć delikatną poprawę w czasie wyszukiwania. Jednakże przy podziale na fragmenty nie jest sprawdzane, czy na ich łączeniach nie znajdują się wzorce, zatem najprawdopodobniej czasy te byłyby zbliżone do zmierzonych czasów bez podziału na fragmenty.