AKWB zadanie 4

1. Opis algorytmu

1.1 Wczytanie instancji:

Program prosi użytkownika o podanie nazwy pliku zawierającego instancje wejściową. Instancja przedstawia multizbiór długości fragmentów uzyskanych w wyniku trawienia cząsteczki DNA enzymem restrykcyjnym. Długości te oddzielone są spacją. Instancja wczytywana jest do wektora.

1.2 Sprawdzenie liczności zbioru i liczby cięć enzymem restrykcyjnym.

Zależność pomiędzy liczbą cięć (k) enzymem a licznością multizbioru A (|A|):

Wzór ten można sprowadzić do postaci:

Z tego możemy wyznaczyć już k i obliczyć liczbę cięć

Odrzucam k2 ponieważ liczba cięć restrykcyjnych musi być większa od 0. Zatem przechodzę do obliczenia k1:

Na tej podstawie sprawdzamy czy istnieje liczba całkowita k z której można wygenerować multizbiór A o tej samej liczności jak ten podany na wejściu. Jeśli nie jest to możliwe to oznacza błąd w instancji. Wypisywany jest komunikat o błędzie i praca programu zostaje przerwana.

1.3 Poszukiwanie mapy

Elementy znalezionej mapy restrykcyjnej będą przechowywane w wektorze globalnym restriction\_map. Pierwszym elementem mapy jest element będący różnicą dwóch największych elementów multizbioru. Elementy instancji nie muszą być posortowane, więc największe elementy niekoniecznie będą znajdować się na końcu multizbioru. W celu znalezienia największego elementu następuje iteracja po wektorze z instancją i porównywane są ze sobą kolejne wartości elementów. Kolejna iteracja służy znalezieniu drugiego największego elementu. Wartości elementów są również ze sobą porównywane, ale dodatkowym warunkiem jest to, że drugi największy element musi być mniejszy od największego elementu. Następnie liczy ich różnicę, która stanowi długość pierwszego elementu mapy restrykcyjnej. Program sprawdza czy dany element znajduje się w multizbiorze. Jeśli tak, to element ten jest umieszczany w wektorze restriction\_map i wywołana jest funkcja find\_map. Jeśli różnica nie została znaleziona w multizbiorze, na ekran wypisywany jest komunikat informujący o błędzie w instancji i program przerywa działanie. Funkcja find\_map przyjmuje dwa parametry, jednym z nich jest obecna liczba elementów mapy a drugim największy element multizbioru. Wewnątrz funkcji znajduje się wektor, do którego dodawane są wykorzystane elementy multizbioru oraz elementy mapy. Dla każdego elementu nie będącego pierwszym elementem mapy liczona jest również odległość między innymi cięciami oraz odległość od „lewego końca” cząsteczki. Odległości te również umieszczane są w wektorze przechowującym zużyte elementy. Pod uwagę nie są brane odległości od „prawego końca”, ponieważ będą one stanowiły odległość ostatniego elementu mapy od poszczególnych cięć. Uwzględnienie odległości od „prawego końca” cząsteczki spowodowałoby zdublowanie tych wartości, przez co algorytm nie mógłby znaleźć mapy restrykcyjnej danej cząsteczki. Algorytm porównuje dotychczasowy rozmiar mapy podany jako parametr funkcji z maksymalnym rozmiarem mapy zapisanym w globalnej zmiennej maxind, która jest większa o jeden od wyliczonej wcześniej liczby cięć cząsteczki. Jeśli rozmiar mapy jest równy zmiennej maxind, mapa została znaleziona i funkcja zwraca wartość true. Jeśli rozmiar mapy nie jest równy maxind, oznacza to, że mapa jest jeszcze niekompletna. Każdy element multizbioru jest potencjalnym kandydatem na kolejny element mapy. Jeśli dany element przyjmuje taką samą wartość jak usunięty wcześniej element z mapy, algorytm przechodzi do kolejnej iteracji i co za tym idzie, kolejnego potencjalnego kandydata. W celu sprawdzenia czy dany element nie został już zużyty, następuje zliczenie występowania elementów o tej wartości w wektorach instancji i tym z wykorzystanymi elementami. Jeśli ilość wystąpień w multizbiorze jest większa niż ilość wystąpień w wektorze elementów zużytych, wartość tego elementu jest przypisana do zmiennej candidate. Następnie liczone są odległości kandydata od „lewego” i „prawego” końca cząsteczki oraz odległości między cięciami, gdyby kolejnym elementem mapy był ten kandydat. Liczona jest ilość występowań elementów o danych długościach w wektorach multisetA i used. Jeśli dodawany jest ostatni element mapy, to by kandydat mógł nim być, odległość kandydata od „prawego” końca musi być równa zero, odległość od „lewego” końca musi być równa długości całej cząsteczki i wszystkie odległości między cięciami muszą być obecne w multizbiorze i niezużyte. Przy dodawaniu elementów niebędących ostatnim elementem, by kandydat został dodany do mapy, to wszystkie wyliczone wcześniej odległości muszą występować w multizbiorze i ilość elementów o tej wartości w multizbiorze musi być większa od ich ilości w wektorze z zużytymi elementami. Jeżeli kandydat spełnia wyżej wymienione warunki, zostaje dodany na koniec wektora restriction\_map i następuje kolejne wywołanie funkcji find\_map z zaktualizowanym parametrem dotychczasowej liczby elementów mapy i niezmienionym parametrem oznaczającym całkowitą długość cząsteczki. Jeśli funkcja zwróci wartość true, to to wywołanie funkcji również zwróci true. W przeciwnym wypadku zdejmowany jest z mapy ostatni element i jego wartość zapisana jest w zmiennej deleted. Algorytm przechodzi do kolejnej iteracji poszukującej potencjalnych kandydatów na element mapy. Jeśli żaden z elementów multizbioru nie spełni warunków bycia kolejnym elementem mapy, funkcja zwraca wartość false.

1.4 Obliczenie czasu pracy algorytmu.

Czas pracy algorytmu liczony jest od momentu wczytania instancji z pliku tekstowego do momentu zwrócenia mapy restrykcyjnej. Czas pracy obliczany jest w minutach, sekundach i mikrosekundach. Jeśli obliczenia trwają dłużej niż 60 minut to wyświetlany zostaje komunikat o zbyt długim czasie obliczeń.

2. Testy na własnych instancjach

a) nieprawidłowe wejście

Podaj nazwe pliku: 01

Wczytana instancja: 1 3 6 7 4 11 13 5

Dlugosc instancji: 8

Niepoprawna ilosc elementow w multizbiorze.

b)

Podaj nazwe pliku: 5\_ciec

Wczytana instancja: 192 114 55 82 37 128 64 137 187 256 64 219 27 73 251 32 119 59 155 5 192

Dlugosc instancji: 21

Przewidywana liczba ciec restrykcyjnych: 5

Mapa: 5 32 27 55 73 64

Czas pracy: 627 MicroSeconds

Podaj nazwe pliku: 5b\_ciec

Wczytana instancja: 46 19 8 14 7 21 25 38 32 59 13 52 6 13 57 5 21 11 27 2 34

Dlugosc instancji: 21

Przewidywana liczba ciec restrykcyjnych: 5

Mapa: 2 5 6 8 13 25

Czas pracy: 559 MicroSeconds

c)

Podaj nazwe pliku: 8\_ciec

Wczytana instancja: 137 269 84 309 393 334 56 100 139 239 70 170 33 44 257 276 107 58 114 346 418 118 232 162 161 37 195 95 37 207 62 279 360 70 430 12 195 77 323 139 151 25 199 223 81

Dlugosc instancji: 45

Przewidywana liczba ciec restrykcyjnych: 8

Mapa: 12 25 33 37 44 56 62 77 84

Czas pracy: 945 MicroSeconds

Podaj nazwe pliku: 8b\_ciec

Wczytana instancja: 104 236 4 161 165 226 28 82 179 101 118 142 38 54 221 123 140 103 114 241 230 42 156 96 9 22 207 125 80 222 14 51 127 60 245 15 47 5 105 19 194 65 118 23 76

Dlugosc instancji: 45

Mapa: 4 5 14 28 54 22 38 65 15

Czas pracy: 164 MicroSeconds

d)

Podaj nazwe pliku: 11a\_ciec

Wczytana instancja: 37 26 13 26 28 22 24 5 17 41 18 35 6 20 24 20 25 5 15 9 30 35 22 9 6 7 20 15 4 22 4 4 16 13 32 2 11 4 11 8 19 33 16 39 9 8 3 30 18 22 10 31 17 19 15 31 17 7 29 13 22 28 6 9 37 10 2 2 39 15 7 2 2 3 13 12 9 13

Dlugosc instancji: 78

Przewidywana liczba ciec restrykcyjnych: 11

Znaleziona mapa: 2 4 3 2 4 4 3 2 7 6 2 2

Czas pracy: 36 Minutes, 48 Seconds, 946 MicroSeconds

Podaj nazwe pliku: 11b\_ciec

Wczytana instancja: 3 5 8 17 22 22 25 32 39 44 45 47 55 67 68 77 77 81 84 87 89 92 95 99 101 116 121 123 124 132 145 154 155 158 171 176 176 179 196 200 200 222 226 232 239 244 247 253 277 277 281 299 313 316 321 321 324 354 358 376 380 397 402 405 408 422 453 475 477 492 497 500 509 554 576 593 598 601

Dlugosc instancji: 78

Przewidywana liczba ciec restrykcyjnych: 11

Znaleziona mapa: 3 5 17 22 45 32 55 68 77 81 95 101

Czas pracy: 293 MicroSeconds

e)

Podaj nazwe pliku: 14\_ciec

Wczytana instancja: 385 383 152 260 133 306 341 79 144 459 251 350 547 271 367 468 273 35 79 113 337 369 198 154 74 22 188 109 74 140 75 55 216 20 346 9 132 97 125 555 153 55 176 424 57 292 23 334 123 250 96 295 118 110 187 363 446 78 282 460 556 359 218 490 415 438 405 564 207 228 119 361 35 44 193 101 482 403 187 470 262 272 272 450 328 87 8 85 65 360 197 210 62 181 284 145 105 195 325 216 149 172 97 22 306 222 77 35 479 44 90 12 326 263 90 227 238 411 137 230

Dlugosc instancji: 120

Przewidywana liczba ciec restrykcyjnych: 14

Mapa: 8 77 20 35 55 23 12 62 75 44 35 22 22 65 9

Czas pracy: 2 Minutes, 7 Seconds, 527 MicroSeconds

Podaj nazwe pliku: 14b\_ciec

Wczytana instancja: 3 5 7 8 12 14 15 20 21 25 26 26 29 34 37 41 45 46 46 49 51 55 58 59 63 66 66 71 71 73 74 83 85 88 88 92 97 100 101 108 109 113 122 124 125 129 134 134 137 138 139 154 159 161 164 168 175 179 180 183 189 196 197 209 213 222 223 225 227 230 235 238 247 252 262 267 281 285 286 288 288 293 296 298 313 333 335 340 347 352 354 359 361 362 386 386 406 410 420 423 427 432 435 449 465 474 494 508 511 515 520 523 548 574 599 619 633 640 645 648

Dlugosc instancji: 120

Przewidywana liczba ciec: 14

Mapa: 3 5 7 14 20 25 26 37 46 55 58 66 73 88 125

Czas pracy: 747 MicroSeconds

f)

Podaj nazwe pliku: 11\_ciec

Wczytana instancja: 2 2 2 3 3 3 3 3 3 3 5 5 5 5 5 5 2 2 2 2 7 7 7 9 9 9 9 15 15 15 7 7 7 3 3 3 15 15 15 15 3 2 3 2 4 7 7 7 19 19 15 15 15 22 25 25 25 20 22 22 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 15 22 22

Dlugosc instancji: 78

Przewidywana liczba ciec restrykcyjnych: 11

Nie znaleziono rozwiazan

3. Testy na gotowych instancjach

**ins-PDP-11a-asc**

Dlugosc instancji: 78

Przewidywana liczba ciec restrykcyjnych: 11

Mapa: 4 6 5 8 3 9 5 2 4 7 8 6

Czas pracy: 2 Seconds, 518 MicroSeconds

**ins-PDP-11b-asc**

Dlugosc instancji: 78

Przewidywana liczba ciec restrykcyjnych: 11

Mapa: 38 74 27 66 42 15 89 47 35 13 12 54

Czas pracy: 922 MicroSeconds

**ins-PDP-11b-desc**

Dlugosc instancji: 78

Przewidywana liczba ciec restrykcyjnych: 11

Mapa: 38 74 27 66 42 15 89 47 35 13 12 54

Czas pracy: 3 Seconds, 94 MicroSeconds

**ins-PDP-12a-asc**

Dlugosc instancji: 91

Przewidywana liczba ciec restrykcyjnych: 12

Mapa: 4 6 5 8 3 9 5 2 4 7 8 6 6

Czas pracy: 7 Seconds, 91 MicroSeconds

**ins-PDP-13a-asc**

Dlugosc instancji: 105

Przewidywana liczba ciec restrykcyjnych: 13

Mapa: 3 6 6 8 7 4 2 5 9 3 8 5 6 4

Czas pracy: 9 Seconds, 234 MicroSeconds

**ins-PDP-14a-asc**

Dlugosc instancji: 120

Przewidywana liczba ciec restrykcyjnych: 14

Mapa: 4 6 5 8 3 9 5 2 4 7 8 6 6 3 5

Czas pracy: 58 Seconds, 475 MicroSeconds

**ins-PDP-14b-asc**

Dlugosc instancji: 120

Przewidywana liczba ciec restrykcyjnych: 14

Mapa: 54 12 13 35 47 89 15 42 66 27 74 38 25 57 79

Czas pracy: 215 MicroSeconds

**ins-PDP-14b-desc**

Dlugosc instancji: 120

Przewidywana liczba ciec restrykcyjnych: 14

Mapa: 54 12 13 35 47 89 15 42 66 27 74 38 25 57 79

Czas pracy: 22 Seconds, 240 MicroSeconds

4. Wnioski.

Mapa w sekwencjach posortowanych znajdowana jest szybciej niż w tych ułożonych losowo. Sekwencje posegregowane w sposób malejący mają dłuższy czas poszukiwania mapy niż sekwencje ułożone rosnąco.