САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ ФАКУЛЬТЕТ ИНФОКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Отчет по лабораторной работе №2 по курсу «Алгоритмы и структуры данных» Тема: Сортировка слиянием. Метод декомпозиции.

Выполнил: Криличевский М. Е.

Номер группы: К3139

Проверил: Афанасьев А. В.

Санкт-Петербург

2024 г.

Содержание

Задание 1	3
Задание 3	7
Задание 4	9
Задание 5	12
Задание 6	15
Задание 9	19
Вывод	24

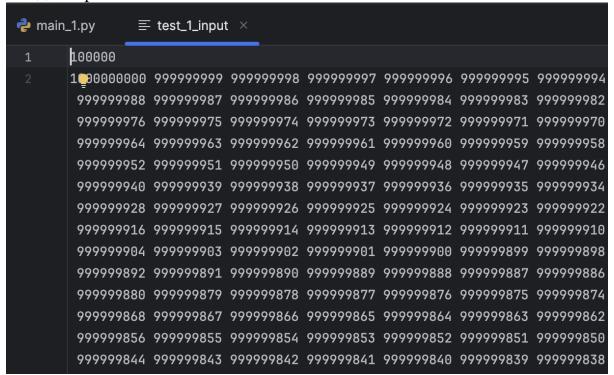
1 задача. Сортировка слиянием

- 1. Используя *псевдокод* процедур Merge и Merge-sort из презентации к Лекции 2 (страницы 6-7), напишите программу сортировки слиянием на Python и проверьте сортировку, создав несколько рандомных массивов, подходящих под параметры:
 - Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 2 \cdot 10^4$) число элементов в массиве. Во второй строке находятся n различных целых чисел, по модулю не превосходящих 10^9 .
 - Формат выходного файла (output.txt). Одна строка выходного файла с отсортированным массивом. Между любыми двумя числами должен стоять ровно один пробел.
 - Ограничение по времени. 2сек.
 - Ограничение по памяти. 256 мб.
- 2. Для проверки можно выбрать наихудший случай, когда сортируется массив размера 1000, 10^4 , 10^5 чисел порядка 10^9 , отсортированных в обратном порядке; наилучший, когда массив уже отсортирован, и средний. Сравните, например, с сортировкой вставкой на этих же данных.
- 3. Перепишите процедуру Merge так, чтобы в ней не использовались сигнальные значения. Сигналом к остановке должен служить тот факт, что все элементы массива L или R скопированы обратно в массив A, после чего в этот массив копируются элементы, оставшиеся в непустом массиве.
- *или* перепишите процедуру Merge (и, соответственно, Merge-sort) так, чтобы в ней не использовались значения границ и середины p, r и q.

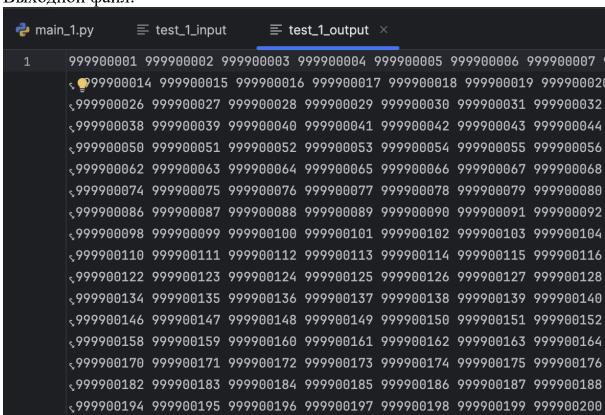
```
Решение:
from lab_2.utils import *
def merge(left_list, right_list):
     res = []
     i = j = 0
     left_len, right_len = len(left_list), len(right_list)
     for _ in range(left_len + right_len):
   if i < left_len and j < right_len:
        if left_list[i] <= right_list[j]:</pre>
                     res.append(left_list[i])
                else:
                     res.append(right_list[i])
          j += 1
elif i == left_len:
    res.append(right_list[j])
          j += 1
elif j == right_len:
                res.append(left_list[i])
                i += 1
     return res
def merge_sort(x):
     if \overline{len}(x) \ll 1:
          return x
     median_x = len(x) // 2
     left_list = merge_sort(x[:median_x])
     right_list = merge_sort(x[median_x:])
return merge(left_list, right_list)
if ___name__=="__main__":
     INPUT_FILE = "../txtf/test_1_input"
OUTPUT_FILE = "../txtf/test_1_output"
     data = read_input(INPUT_FILE)
     if data:
          sorted_data = merge_sort(data)
          write_output(output_file, sorted_data)
print(first_check(input_file_n(INPUT_FILE),operation_with_file(INPUT_FILE))
_FILE)))
```

Объяснение решения: с помощью функции Merge () берутся два отсортированных списка и объединяются в один список: с помощью цикла for на каждом шагу отбирается меньший элемент из двух списков и добавляется в итоговый список. Функция Merge sort () работает по принципу «разделяй и властвуй»: исходный список делится на два подсписка, которые продолжают делиться, пока не будут отсортированы. Дальше начиная с самых маленьких подсписков происходит объединение данных подсписков пока не будет достигнут итоговый отсортированный массив.

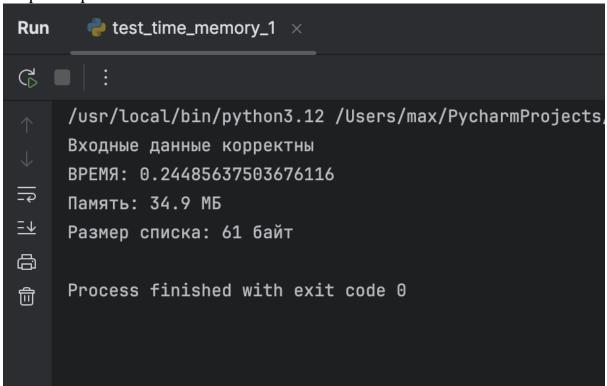
Входной файл:



Выходной файл:



Затраты времени и памяти:



3 задача. Число инверсий

Инверсией в последовательности чисел A называется такая ситуация, когда i < j, а $A_i > A_j$. Количество инверсий в последовательности в некотором роде определяет, насколько близка данная последовательность к отсортированной. Например, в сортированном массиве число инверсий равно 0, а в массиве, сортированном наоборот - каждые два элемента будут составлять инверсию (всего n(n-1)/2).

Дан массив целых чисел. Ваша задача — подсчитать число инверсий в нем. Подсказка: чтобы сделать это быстрее, можно воспользоваться модификацией сортировки слиянием.

- Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 10^5$) число элементов в массиве. Во второй строке находятся n различных целых чисел, по модулю не превосходящих 10^9 .
- **Формат выходного файла (output.txt).** В выходной файл надо вывести число инверсий в массиве.
- Ограничение по времени. 2сек.
- Ограничение по памяти. 256 мб.

Решение:

```
from lab_2.utils import *

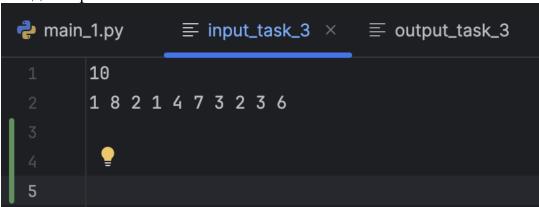
def inverse(arr):
    count = 0
    n = len(arr)
    for i in range(n):
        if arr[i] > arr[j]:
            count += 1
    return count

if __name__ == "__main__":
    FILE_INPUT = "../txtf/input_task_3"
    FILE_OUTPUT = "../txtf/output_task_3"
    file = operation_with_file(FILE_INPUT)
    res = inverse(file)
    check_list =

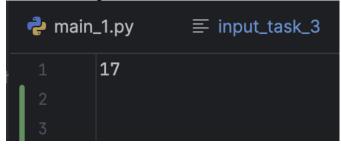
check(input_file_n(FILE_INPUT),operation_with_file(FILE_INPUT))
    if check_list:
        output = output_file(FILE_OUTPUT, str(res))
        print("Входные данные корректны")
    else:
        output_file(FILE_OUTPUT, "Ошибка входных данных")
        print("Ошибка входных данных")
```

Объяснение решения: с помощью функции inverse () мы перебираем пары элементов списка, поступающего на вход как аргумент функции. Если первый элемент пары больше второго и индекс первого элемента пары меньше индекса второго, то мы увеличиваем счетчик на единицу, в результате возвращаем его.

Входной файл:



Выходной файл:



Затраты времени и памяти:

```
/usr/local/bin/python3.12 /Users/max/Pycha
Входные данные корректны
ВРЕМЯ: 0.00046033399121370167
Память: 9.9 МБ
Размер списка: 184 байт

Process finished with exit code 0
```

4 задача. Бинарный поиск

В этой задаче вы реализуете алгоритм бинарного поиска, который позволяет очень эффективно искать (даже в огромных) списках при условии, что список отсортирован. Цель - реализация алгоритма двоичного (бинарного) поиска.

- Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 10^5$) число элементов в массиве, и последовательность $a_0 < a_1 < ... < a_{n-1}$ из n различных положительных целых чисел в порядке возрастания, $1 \le a_i \le 10^9$ для всех $0 \le i < n$. Следующая строка содержит число k, $1 \le k \le 10^5$ и k положительных целых чисел $b_0, ... b_{k-1}, 1 \le b_j \le 10^9$ для всех $0 \le j < k$.
- Формат выходного файла (output.txt). Для всех i от 0 до k-1 вывести индекс $0 \le j \le n-1$, такой что $a_i = b_j$ или -1, если такого числа в массиве нет.
- Ограничение по времени. 2сек.
- Ограничение по памяти. 256 мб.
- Пример:

input.txt	output.txt
5	20-10-1
1 5 8 12 13	
5	
8 1 23 1 11	

В этом примере есть возрастающая последовательность из $a_0=1, a_1=5, a_2=8, a_3=12$ и $a_4=13$ длиной в n=5 и пять чисел для поиска: 8 1 23 1 11. Видно, что $a_2=8$ и $a_0=1$, но чисел 23 и 11 нет в последовательности a, поэтому они имеют индекс -1. В итоге ответ: 2 0 -1 0 -1.

```
Решение:
from lab_2.utils import *
from lab_2.utils import
second_file_operation_list1,second_file_operation_list2,second_check
 ,second_file_operation_line1,second_file_operation_line3,
output_file, string
def binary_search(arr, target):
    left, right = 0, len(arr) - 1
    while left <= right:
        mid = (left + right) // 2
        if arr[mid] == target:</pre>
                       return mid
               elif arr[mid] < target:
   left = mid + 1</pre>
                       right = mid - 1
        return -1
def answer(a, b):
        return string([binary_search(a, index) for index in b])
if __name__=="__main__":
       __name__ == "__main__":

FILE_INPUT = "../txtf/input_task_4"

FILE_OUTPUT = "../txtf/output_task_4"

file_1 = second_file_operation_list1(FILE_INPUT)

file_2 = second_file_operation_list2(FILE_INPUT)

line1 = second_file_operation_line1(FILE_INPUT)

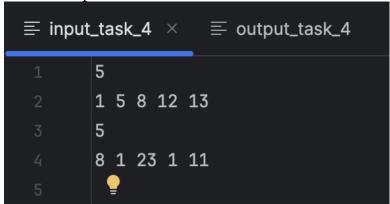
line3 = second_file_operation_line3(FILE_INPUT)

check = second_check(line1, line3, file_1, file_2)
       if check:
               results = answer(file_1, file_2)
               file_o = output_file(FILE_OUTPUT, results)
print("Входные данные корректны")
        else:
               file_o = output_file(FILE_OUTPUT, "Ошибка входных данных")
               print("Ошибка входных данных")
```

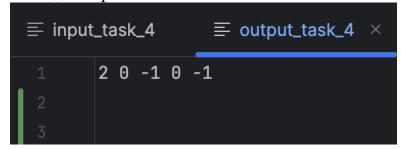
Объяснение решения:

Данный код реализует поиск индексов элементов из массива 'b' в массиве 'a' с помощью бинарного поиска. Программа читает два массива из входного файла, используя функции из 'lab_2. utils', проверяет корректность данных и, если они валидны, выполняет поиск индексов каждого элемента из второго массива в первом. Результаты записываются в строку и сохраняются в выходной файл. Если данные некорректны, в файл записывается сообщение об ошибке. Алгоритм бинарного поиска используется, так как массив 'a' предполагается отсортированным, обеспечивая эффективность для каждого элемента.

Входной файл:



Выходной файл:



Затраты времени и памяти:

```
/usr/local/bin/python3.12 /Users/max/Pycharm/
Входные данные корректны
ВРЕМЯ: 0.000189124999451451
Память: 9.6 МБ
Размер списка №1: 120 байт
Размер списка №2: 120 байт

Process finished with exit code 0
```

Правило большинства — это когда выбирается элемент, имеющий больше половины голосов. Допустим, есть последовательность А элементов a1, a2, ...an, и нужно проверить, содержит ли она элемент, который появляется больше, чем n/2 раз. Наивный метод как это сделать:

```
Majority(A):
for i from 1 to n:
    current_element = a[i]
    count = 0
    for j from 1 to n:
        if a[j] = current_element:
            count = count+1
    if count > n/2:
        return a[i]
return "нет элемента большинства"
```

Очевидно, время выполнения этого алгоритма квадратично. Ваша цель - использовать метод "Разделяй и властвуй" для разработки алгоритма проверки, содержится ли во входной последовательности элемент, который встречается больше половины раз, за время $O(n\log n)$.

- Формат входного файла (input.txt). В первой строке входного файла содержится число n ($1 \le n \le 10^5$) число элементов в массиве. Во второй строке находятся n положительных целых чисел, по модулю не превосходящих 10^9 , $0 \le a_i \le 10^9$.
- Формат выходного файла (output.txt). Выведите 1, если во входной последовательности есть элемент, который встречается строго больше половины раз; в противном случае 0.
- Ограничение по времени. 2сек.
- Ограничение по памяти. 256 мб.
- Пример 1:

input.txt	output.txt
5	1
23922	

Число "2"встречается больше 5/2 раз.

• Пример 2:

input.txt	output.txt
4	0
1234	

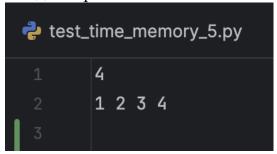
Нет элемента, встречающегося больше n/2 раз.

Решение:

Объяснение решения:

Этот код ищет элемент, который встречается более половины раз от общей длины массива. Выполняется проверка каждого элемента массива: для каждого из них подсчитывается, сколько раз он встречается в массиве. Данные для анализа считываются из входного файла, их корректность проверяется с помощью функций из модуля `lab_2.utils`. В случае успешной проверки строковое значение единицы или нуля записывается в выходной файл. Если входные данные некорректны — в выходной файл указывается сообщение об ощибке.

Входной файл:



Выходной файл:



Затраты времени и памяти:

```
/usr/local/bin/python3.12 /Users/max/Pycharm/
Входные данные корректны
ВРЕМЯ: 0.0004195420042378828
Память: 10.1 МБ
Размер списка: 88 байт

Process finished with exit code 0
```

Используя псевдокод процедур Find Maximum Subarrayu Find Max Crossing Subarray из презентации к Лекции 2 (страницы 25-26), напишите программу по- иска максимального подмассива.

Примените ваш алгоритм для ответа на следующий вопрос. Допустим, у нас есть данные по акциям какой-либо фирмы за последний месяц (год, или иной срок).

Проанализируйте этот срок и выдайте ответ, в какой из дней при покупке единицы акции данной фирмы, и в какой из дней продажи, вы бы получили максимальную прибыль? Выдайте дату покупки, дату продажи и максимальную прибыль.

Вы можете использовать любые данные для своего анализа. Например, я на- брала в Google "акции" и мне поиск выдал акции Газпрома, тут - можно скачать информацию по стоимости акций за любой период. (Перейдя по ссылке, нажмите на вкладку "Настройки"→ "Скачать") Соответственно, вам нужно только выбрать данные, посчитать изменение це- ны и применить алгоритм поиска максимального подмассива.

- Формат входного файла в данном случае на ваше усмотрение.
- Формат выходного файла (output.txt). Выведите название фирмы, рассматриваемый вами срок изменения акций, дату покупки и дату продажи единицы акции, чтобы получилась максимальная выгода; и сумма этой при- были.

Решение:

```
def max_subarray(prices, dates):
     if not prices or len(prices) != len(dates):
           return [0, 0, 0, None, None]
     min_price = float('inf')
     min_price_index = -1
     max\_profit = 0
     for i in range(len(prices)):
           if prices[i] < min_price:</pre>
                min_price = prices[i]
                min_price_index = i
          potential_profit = prices[i] - min_price
if potential_profit > max_profit and i > min_price_index :
    max_profit = potential_profit
                best_buy_index = min_price_index
best_sell_index = i
     if max_profit > 0 and best_buy_index !=-1 and best_sell_index !=
-1:
           best_buy_price = prices[best_buy_index]
best_sell_price = prices[best_sell_index]
best_buy_date = dates[best_buy_index]
best_sell_date = dates[best_sell_index]
return [best_buy_price, best_sell_price, max_profit,
best_buy_date, best_sell_date]
           return [0, 0, 0, None, None]
if __name__=="__main__":
```

Объяснение решения:

Этот код рассчитывает наибольшую возможную прибыль от покупки и продажи акций, основываясь на списке цен и соответствующих дат. Программа находит такие значения, чтобы покупка происходила по минимальной цене, а продажа — по максимальной цене, следующей за этой минимальной. Результат возвращается в виде цены покупки, цены продажи, прибыли, даты покупки, даты продажи.

Входной файл:

≡ input	_task_6 ×	≡ output_task_6
1	Date	Price
2	27.09.2024	4024.0
3	30.09.2024	4007.5
4	01.10.2024	3963.0
5	02.10.2024	3907.5
6	03.10.2024	3969.0
7	04.10.2024	3945.5
8	07.10.2024	3928.5
9	08.10.2024	3982.5
10	09.10.2024	3933.5
11	10.10.2024	3979.5
12	11.10.2024	3975.0
13	14.10.2024	4074.0
14	15.10.2024	4091.0
15	16.10.2024	4057.0
16	17.10.2024	3983.5
17	18.10.2024	3968.5
18	21.10.2024	4031.0
19	22.10.2024	3964.5
20	23.10.2024	3883.5
21	24.10.2024	3969.5
22	25.10.2024	3834.5

Выходной файл:

Затраты времени и памяти:

/usr/local/bin/python3.12 /Users/max/Pycharm

ВРЕМЯ: 0.00028433398983906955

Память: 9.6 МБ

Process finished with exit code 0

9 задача. Метод Штрассена для умножения матриц

Умножение матриц. Простой метод. Если есть квадратные матрицы $X=(x_{ij})$ и $Y=(y_{ij})$, то их произведение $Z=X\cdot Y\Rightarrow z_{ij}=\sum_{k=1}^n x_{ik}\cdot y_{kj}$. Нужно вычислить n^2 элементов матрицы, каждый из которых представляет собой сумму n значений.

```
Matrix_Multiply(X, Y)::
    n = X.rows
Z - квадратная матрица размера n
    for i = 1 to n:
        for j = 1 to n:
        z[i,j] = 0
        for k = 1 to n:
        z[i,j] = z[i,j] + x[i,k]*y[k,j]
    return Z
```

Задачу умножения матриц достаточно легко разбить на подзадачи, поскольку произведение можно составлять из блоков. Разобьём каждую из матриц X и Y на четыре блока размера $\frac{n}{2} \times \frac{n}{2}$:

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \ Y = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

Тогда их произведение выражается в терминах этих блоков по обычной формуле умножения матриц:

$$XY = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

Вычислив рекурсивно восемь произведений AE, BG, AF, BH, CE, DG, CF, DH и просуммировав их за время $O(n^2)$, мы вычислим необходимое нам произведение матриц. Соответствующее рекуррентное соотношение на время работы алгоритма

$$T(n) = 8T(\frac{n}{2}) + O(n^2).$$

Какое получилось время у предыдущего рекурсивного алгоритма? Да, ничуть не лучше наивного. Однако его можно ускорить с помощью алгебраического трюка: для вычисления произведения XY достаточно перемножить cemb пар матриц размера $\frac{n}{2} \times \frac{n}{2}$, после чего хитрым образом (u как только Штрассен догадался?) получить ответ:

$$XY = \begin{bmatrix} P_5 + P_4 - P_2 + P_6 & P_1 + P_2 \\ P_3 + P_4 & P_1 + P_5 - P_3 - P_7 \end{bmatrix}$$

где

$$P_1 = A(F - H),$$
 $P_5 = (A + D)(E + H),$
 $P_2 = (A + B)H,$ $P_6 = (B - D)(G + H),$
 $P_3 = (C + D)E,$ $P_7 = (A - C)(E + F),$
 $P_4 = D(G - E).$

- **Цель**. Применить метод Штрассена для умножения матриц и сравнить его с простым методом. *Найти размер матриц п, при котором метод Штрассена работает существенно быстрее простого метода*.
- Формат входа. Стандартный ввод или input.txt. Первая строка размер квадратных матриц n для умножения. Следующие строки соответсвенно сами значения матриц A и B.
- Формат выхода. Стандартный вывод или output.txt. Матрица $C = A \cdot B$.

Решение: Multiply:

Strassen:

```
THRESHOLD = 16
def split(M):
     n = len(M) // 2
     return (
    [[M[i][j] for j in range(n)] for i in range(n)], # A11
    [[M[i][j] for j in range(n, n * 2)] for i in range(n)],
A12
           [[M[i][j] for j in range(n)] for i in range(n, n * 2)],
A21
           [[M[i][j] for j in range(n, n * 2)] for i in range(n, n *
2)],
def merge(C11, C12, C21, C22):
     n = len(C11)
     C = [[0] for _ in range(n * 2)] for _ in range(n * 2)]
     for i in range(n):
          for j in range(n):
    C[i][j] = C11[i][j]
                C[i][j + n] = C12[i][j]
C[i + n][j] = C21[i][j]
C[i + n][j + n] = C22[i][j]
     return C
def strassen(A, B):
     n = len(A)
     if n <= THRESHOLD:</pre>
           return [[sum(A[i][k] * B[k][j] for k in range(n)) for j in
range(n)] for i in range(n)]
          A11, A12, A21, A22 = split(A)
B11, B12, B21, B22 = split(B)
          P1 = strassen(A11 + A22, B11 + B22)
P2 = strassen(A21 + A22, B11)
          P3 = strassen(A11, B12 - B22)
P4 = strassen(A22, B21 - B11)
          P5 = strassen(A11 + A12, B22)
          P6 = strassen(A21 - A11, B11 + B12)
          P7 = strassen(A12 - A22, B21 + B22)
          C11 = P1 + P4 - P5 + P7
          C12 = P3 + P5
          C21 = P2 + P4
          C22 = P1 - P2 + P3 + P6
           return merge(C11, C12, C21, C22)
def Check(file_inp, file_out):
    with open(file_inp, 'r') as f:
        n = int(f.readline())
          A = [list(map(int, f.readline().split())) for k in range(n)]
```

```
B = [list(map(int, f.readline().split())) for k in range(n)]
C = strassen(A, B)
with open(file_out, 'w') as f:
    for res1 in C:
        f.write(' '.join(map(str, res1)) + '\n')

if __name__ == "__main__":
    Check("../txtf/input_task_9", "../txtf/output_task_9")
```

Объяснение решения:

Этот перемножение квадратных код выполняет матриц использованием двух подходов: стандартного метода и ускоренного Штрассена. алгоритма Если размер матрицы меньше 'THRESHOLD', применяется обычный метод, иначе используется рекурсивный алгоритм Штрассена. Для больших матриц он делит их на 4 подматрицы, вычисляет промежуточные результаты, а затем объединяет их в результирующую матрицу. Чтение данных (размер и значения матриц) производится из входного файла, после чего результат записи сохраняется в выходной файл.

Входной файл:

≡ input	_task_9 ×	≡ output_task_9
1	3	
2	1 2 7	
3	3 4 5	
4	9 10 11	
5	16 8 3	
6	6 9 2	
7	8 3 1	

Выходные файлы:

Multiply:

Strassen:

Затраты времени и памяти:

Multiply:

```
/usr/local/bin/python3.12 /Users/max/Pycharm
ВРЕМЯ: 0.0002573749952716753
Память: 9.5 МБ
Process finished with exit code 0
```

Strassen:

```
/usr/local/bin/python3.12 /Users/max/Pycharm
ВРЕМЯ: 0.0003852079971693456
Память: 10.4 МБ
Process finished with exit code 0
```

Вывод

В данной лабораторной работе познакомился и поработал с новыми алгоритмами сортировки и поиска, а также решил несколько интересных задач: про акции и матрицы.