

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт цифрового развития
Кафедра инфокоммуникаций

ОТЧЕТ
ПО ЛАБОРАТОРНОЙ РАБОТЕ №10
дисциплины «Алгоритмизация»

Выполнил:
Магдаев Даламбек Магомедович
2 курс, группа ИВТ-б-о-22-1,
09.03.01 «Информатика и
вычислительная техника»,
направленность (профиль)
«Программное обеспечение средств
вычислительной техники и
автоматизированных систем», очная
форма обучения

(подпись)

Руководитель практики:
Воронкин Р.А., доцент кафедры
инфокоммуникаций

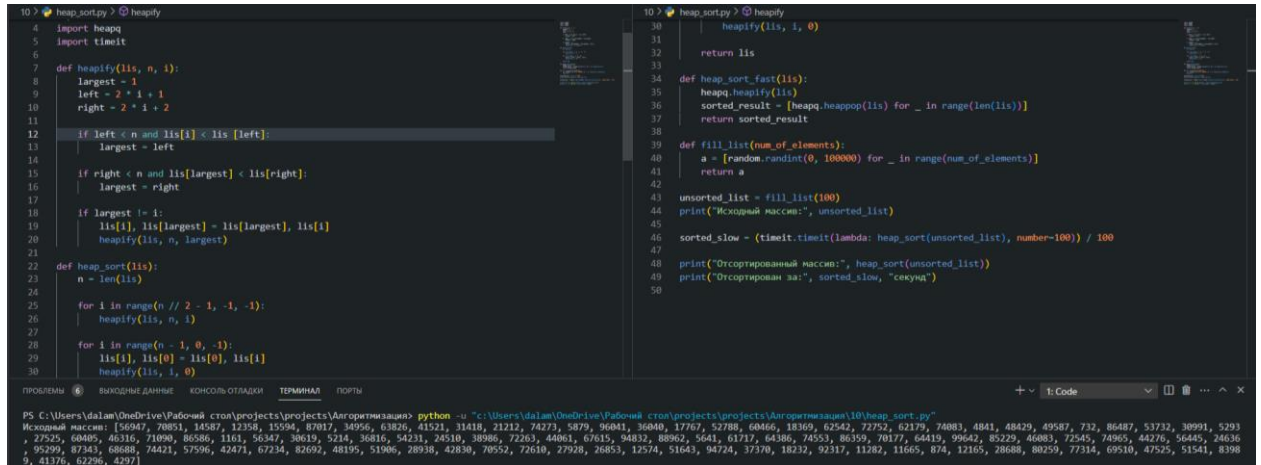
(подпись)

Отчет защищен с оценкой _____ Дата защиты _____

Ставрополь, 2023 г.

Порядок выполнения работы:

1. Написал программу поиска элемента в массиве, автоматического заполнения массива, расчёта тысячи точек, показывающих время поиска элемента в массиве в худшем и среднем случае, вывода графиков, составленных из этих точек, и подсчета корреляции:



```
10 > heap_sort.py > heapify
4 import heapq
5 import timeit
6
7 def heapify(lis, n, i):
8     largest = i
9     left = 2 * i + 1
10    right = 2 * i + 2
11
12    if left < n and lis[i] < lis[left]:
13        largest = left
14
15    if right < n and lis[largest] < lis[right]:
16        largest = right
17
18    if largest != i:
19        lis[i], lis[largest] = lis[largest], lis[i]
20        heapify(lis, n, largest)
21
22 def heap_sort(lis):
23     n = len(lis)
24
25     for i in range(n // 2 - 1, -1, -1):
26         heapify(lis, n, i)
27
28     for i in range(n - 1, 0, -1):
29         lis[i], lis[0] = lis[0], lis[i]
30         heapify(lis, i, 0)
31
32
33 heapify(lis, i, 0)
34
35 return lis
36
37 def heap_sort_fact(lis):
38     heapify(lis)
39     sorted_result = [heapq.heappop(lis) for _ in range(len(lis))]
40     return sorted_result
41
42
43 def fill_list(num_of_elements):
44     a = [random.randint(0, 100000) for _ in range(num_of_elements)]
45     return a
46
47
48 unsorted_list = fill_list(100)
49 print("Исходный массив:", unsorted_list)
50
51 sorted_slow = (timeit.timeit(lambda: heap_sort(unsorted_list), number=100)) / 100
52
53 print("Орсортированный массив:", heap_sort(unsorted_list))
54 print("Орсортирован за:", sorted_slow, "секунда")
55
```

Рис. 1. Код и результат неоптимизированного алгоритма heapsort

Рисунок 1. Код и результат неоптимизированного алгоритма heapsort

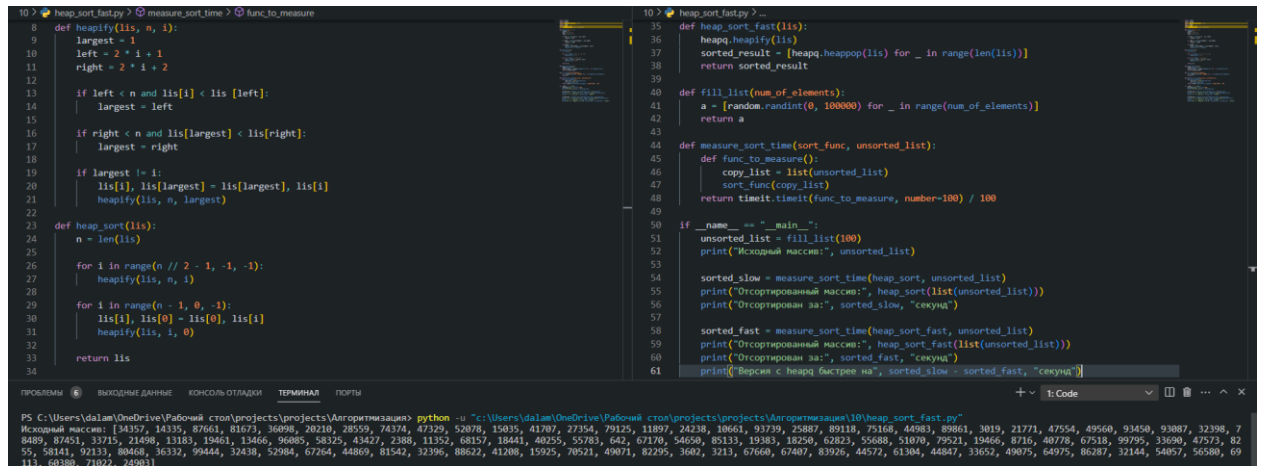
Таблица 1. Сравнение алгоритма Heap Sort с Quick Sort и Merge Sort

Характеристика	Heap Sort	Quick Sort	Merge Sort
Сложность времени	$O(n \log n)$	$O(n^2)$ в худшем случае, $O(n \log n)$ в среднем	$O(n \log n)$
Сложность по памяти	$O(1)$ или $O(\log n)$	$O(\log n)$ в среднем	$O(n)$
Лучший случай	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$
Худший случай	$O(n \log n)$	$O(n^2)$	$O(n \log n)$
Средний случай	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$

Heapsort не требует доп. память, занимает меньше всех места, но считается нестабильным. Quick Sort медленный в худшем случае, занимает больше места, требует доп. память и является нестабильным. Merge Sort стабилен, быстр, не требует доп памяти, но имеет наибольшую сложность по памяти.

2. Произвел оптимизацию алгоритма при помощи встроенной

библиотеки heapq:



```
10 def heap_sort_slow.py > measure_sort_time > func_to_measure
8 def heapify(lis, n, i):
9     largest = i
10    left = 2 * i + 1
11    right = 2 * i + 2
12
13    if left < n and lis[i] < lis[left]:
14        largest = left
15
16    if right < n and lis[largest] < lis[right]:
17        largest = right
18
19    if largest != i:
20        lis[i], lis[largest] = lis[largest], lis[i]
21        heapify(lis, n, largest)
22
23 def heap_sort(lis):
24     n = len(lis)
25
26     for i in range(n // 2 - 1, -1, -1):
27         heapify(lis, n, i)
28
29     for i in range(n - 1, 0, -1):
30         lis[i], lis[0] = lis[0], lis[i]
31         heapify(lis, i, 0)
32
33     return lis
34
10 def heap_sort_fast.py >
35 def heap_sort_fast(lis):
36     heapq.heapify(lis)
37     sorted_result = [heapq.heappop(lis) for _ in range(len(lis))]
38     return sorted_result
39
40 def fill_list(num_of_elements):
41     a = [random.randint(0, 100000) for _ in range(num_of_elements)]
42     return a
43
44 def measure_sort_time(sort_func, unsorted_list):
45     def func_to_measure():
46         copy_list = list(unsorted_list)
47         sort_func(copy_list)
48         return timeit.timeit(func_to_measure, number=100) / 100
49
50 if __name__ == "__main__":
51     unsorted_list = fill_list(100)
52     print("Исходный массив:", unsorted_list)
53
54     sorted_slow = measure_sort_time(heap_sort, unsorted_list)
55     print("Отсортированный массив:", heap_sort(list(unsorted_list)))
56     print("Отсортирован за:", sorted_slow, "секунд")
57
58     sorted_fast = measure_sort_time(heap_sort_fast, unsorted_list)
59     print("Отсортированный массив:", heap_sort_fast(list(unsorted_list)))
60     print("Отсортирован за:", sorted_fast, "секунд")
61
62     print("Версия с heapq быстрее на", sorted_slow - sorted_fast, "секунд")
63
PS C:\Users\dalam\OneDrive\Рабочий стол\projects\projects\Алгоритмизация> python -u "C:\Users\dalam\OneDrive\Рабочий стол\projects\projects\Алгоритмизация\10\heap_sort_fast.py"
Исходный массив: [34357, 14335, 87661, 81673, 36098, 20210, 28559, 74374, 47329, 52078, 15035, 41707, 27354, 79125, 11897, 24238, 10661, 93739, 25887, 89118, 75168, 44983, 89861, 3019, 21771, 47554, 49560, 93450, 93087, 32398, 7
8489, 87451, 33715, 21698, 13183, 19461, 13466, 96805, 50325, 43427, 2388, 11352, 68157, 18441, 40255, 55783, 642, 67170, 54650, 85133, 19383, 18260, 62823, 55688, 51070, 75021, 19466, 8716, 40778, 67518, 99795, 33690, 47573, 82
55, 58141, 92133, 80868, 30332, 39444, 32438, 52984, 67264, 44869, 81542, 32396, 88622, 51206, 15925, 70921, 49071, 82795, 3602, 3313, 67660, 87407, 83926, 44572, 61304, 44847, 33652, 40075, 64075, 86207, 32144, 54057, 36580, 69
113, 60380, 71022, 24903]
```

Рисунок 2. Оптимизированный алгоритм heapsort

3. Применение в реальной жизни:

Системы с ограниченной памятью: Эффективен там, где важно минимизировать использование дополнительной памяти.

Приоритетные очереди: Используется для эффективного управления данными с приоритетом, например в обработке событий.

Потоковая обработка: Подходит для сортировки данных в реальном времени, когда данные постоянно поступают.

Базы данных: Помогает в сортировке больших объемов данных вне основной памяти, через внешнюю сортировку.

Вычислительные задачи с таймингом: Хорош для задач в реальном времени, где важна предсказуемость времени выполнения операций.

Heap Sort выбирают из-за надёжности и предсказуемости, когда стоит избегать риска существенного замедления из-за худшего случая выполнения, как, например, у Quick Sort. Также он полезен, если требуется сортировать данные без дополнительного пространства, например, при выполнении сортировки прямо на физических носителях или в ситуациях с ограниченной доступной памятью.

Анализ сложности:

Добавил рисование графиков зависимости времени сортировки с помощью оптимизированного и не оптимизированного heapsort от количества

элементов в массиве.

Поскольку Heap Sort может выполняться in-place, для его работы теоретически не требуется дополнительное пространство кроме самого массива, который сортируется. При этом методе сортировки:

- Не требуется выделять дополнительный массив для разделения данных.
- Манипуляции с элементами осуществляются в пределах того же массива.
- Требуется ограниченное количество переменных для хранения индексов и временных значений в процессе выполнения алгоритма.

Heap Sort возможно реализовать без рекурсии, при этом алгоритм получается с чистой пространственной сложностью $O(1)$, то есть он будет занимать константное дополнительное пространство, не зависимо от размеров данных. Эта модификация использует только циклы и не требует дополнительной памяти для рекурсивного стека, что делает пространственную сложность чисто константной.

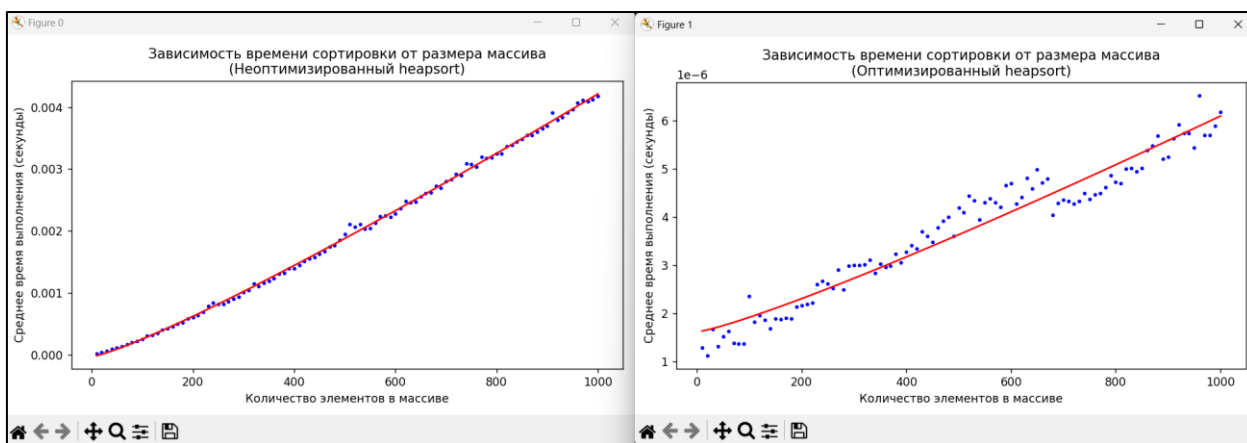


Рисунок 3. Графики зависимости времени сортировки с помощью оптимизированного и не оптимизированного heapsort от количества элементов в массиве

```
10 > task_7.py > ...
11 amount_of_data = 100 # Количество точек
12 and = (amount_of_data - 1) * 10
13 median_time = 0
14 graph_stuff = [1 for i in range(10, and, 10)]
15 xlabel = "Количество элементов в массиве"
16 ylabel = "Среднее время выполнения (секунды)"
17
18 def heapify(lis, n, i):
19     largest = i
20     left = 2 * i + 1
21     right = 2 * i + 2
22
23     if left < n and lis[left] < lis[largest]:
24         largest = left
25
26     if right < n and lis[right] < lis[largest]:
27         largest = right
28
29     if largest != i:
30         lis[i], lis[largest] = lis[largest], lis[i]
31         heapify(lis, n, largest)
32
33 def heap_sort(lis):
34     n = len(lis)
35
36     for i in range(n // 2 - 1, -1, -1):
37         heapify(lis, n, i)
38
39     for i in range(n - 1, 0, -1):
40         lis[i], lis[0] = lis[0], lis[i]
41         heapify(lis, i, 0)
42
43     return lis
44
45 def heap_sort_fast(lis):
46     heapq.heapify(lis)
47     sorted_result = [heapq.heappop(lis) for _ in range(len(lis))]
48     return sorted_result
49
50 def fill_list(num_of_elements):
51     a = [random.randint(0, 100000) for _ in range(num_of_elements)]
52     return a
53
54 def measure_sort_time(sort_func, unsorted_list):
55     def func_to_measure():
56         # Создаем копию, чтобы не влиять на оригинальный список
57         copy_list = list(unsorted_list)
58
59         sort_func(copy_list)
60
61     return timeit.timeit(func_to_measure, number=100) / 100
62
63 def log_n_model(x, a, b):
64     return a * x * np.log(x) + b
65
66 def log_name_time_graph_num():
67     plt.figure(graph_num).set_figwidth(8)
68     plt.title(f"Зависимость времени сортировки от размера массива n (name)")
69     # Зависимость времени сортировки от размера массива n (name)
70     plt.xlabel(xlabel)
71     plt.ylabel(ylabel)
72     plt.tight_layout()
73     plt.grid(False)
74
75     x_data = np.array(graph_stuff)
76     y_data = np.array(list(time.values()))
77
78     params, _ = curve_fit(log_n_model, x_data, y_data)
79
80     a_fit, b_fit = params
81     print(f"Коэффициенты уравнения ((name)): a = {a_fit}, b = {b_fit}")
82
83     x_fit = np.linspace(min(x_data), max(x_data), 100)
84     y_fit = log_n_model(x_fit, *params)
85
86     plt.plot(x_fit, y_fit, "r-", label=f"n log n fit")
87
88     plt.scatter(graph_stuff, time.values(), s=5, c="blue")
89     plt.tight_layout()
90
91 def results(name, func, graph_index):
92     for i in range(10, and, 10):
93         a = fill_list(i)
94         median_time[i] = timeit.timeit(lambda: func(a),
95                                     number=100) / 100
96
97     log(name, median_time, graph_index)
98
99 if __name__ == "__main__":
100     unsorted_list = fill_list(200)
101
102     sorted_slow = measure_sort_time(heap_sort, unsorted_list)
103     print(f"Оценочное значение heap_sort: {a}, sorted_slow: {cex}")
104     sorted_fast = measure_sort_time(heap_sort_fast, unsorted_list)
105     print(f"Оценочное heap_sort < heapq: {a}, sorted_fast: {cex}")
106     print(f"Разрыв в heapq быстрее на", sorted_slow - sorted_fast, "сек")
107
108 PROBLEMY 6 ВЫХОДНЫЕ ДАННЫЕ КОНСОЛЬ ОТЛАДКИ ТЕРМИНАЛ ПОРТЫ
109
110 Оценочное heap_sort с heapq за: 6.825400009242553e-05 сек
111 Разрыв в heapq быстрее на 0.000000000000000000 сек
112 Коэффициенты уравнения (Оценочное значение heap_sort): a = 6.344394891386459e-07, b = -3.196334378577898e-05
113 Коэффициенты уравнения (Оценочное значение heap_sort): a = 6.464182781956654e-10, b = 1.6220775398114602e-06
```

Рисунок 4. Полный код с двумя алгоритмами heapsort и выводом графиков скорости их сортировки

4. Решение задания:

```
10 > task_6.py > ...
1 import heapq
2 import random
3
4
5 def print_ascending_sums(A, B):
6
7     # Сортируем оба массива
8     A.sort()
9     B.sort()
10
11     # min-heap хранит выражения вида (сумма, индекс в A, индекс в B)
12     heap = [(A[i] + B[0], i, 0) for i in range(len(A))]
13     heapq.heapify(heap)
14
15     # Вывод суммы в возрастающем порядке
16     for _ in range(len(A)*2):
17         sum, i, j = heapq.heappop(heap)
18         print(sum, end=' ')
19         if j + 1 < len(B):
20             heapq.heappush(heap, (A[i] + B[j+1], i, j+1))
21
22
23 def fill_list(num_of_elements):
24     a = [random.randint(1, 10) for _ in range(num_of_elements)]
25     return a
26
27
28 n = 5
29 A = fill_list(n)
30 B = fill_list(n)
31 print(sorted(A), "\n", sorted(B))
32 print_ascending_sums(A, B)
33
34 PROBLEMY 6 ВЫХОДНЫЕ ДАННЫЕ КОНСОЛЬ ОТЛАДКИ ТЕРМИНАЛ ПОРТЫ
35
36 PS C:\Users\dalam\OneDrive\Рабочий стол\projects\projects\Алгоритмизация> python -u "c:\Users\dalam\OneDrive\Рабочий
[4, 5, 5, 9, 10]
[2, 3, 5, 8, 10]
6 7 7 7 8 8 9 10 11 12 12 12 13 13 13 14 14 15 15 17 18 19 20
PS C:\Users\dalam\OneDrive\Рабочий стол\projects\projects\Алгоритмизация>
```

Рисунок 3. Код решения задания 6 и результат выполнения

Вывод: в результате выполнения лабораторной работы был изучен алгоритм heap sort и проведено исследование зависимости времени поиска от количества элементов в массиве, показавшее что зависимость время поиска линейно увеличивается с добавлением элементов в массив.