



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

ОТЧЕТ

по лабораторной работе № 3
по курсу «Анализ алгоритмов»
на тему: «Трудоёмкость сортировок»

Студент ИУ7-54Б
(Группа)

(Подпись, дата)

Писаренко Д. П.
(И. О. Фамилия)

Преподаватель

(Подпись, дата)

Волкова Л. Л.
(И. О. Фамилия)

2023 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Аналитический раздел	4
1.1 Алгоритм блочной сортировки	4
1.2 Алгоритм быстрой сортировки	4
1.3 Алгоритм сортировки выбором	4
2 Конструкторский раздел	6
2.1 Требования к программному обеспечению	6
2.2 Описание используемых типов данных	6
2.3 Разработка алгоритмов	7
2.4 Оценка трудоемкости алгоритмов	10
2.4.1 Трудоемкость алгоритма блочной сортировки	10
2.4.2 Трудоемкость алгоритма быстрой сортировки	11
2.4.3 Трудоемкость алгоритма сортировки выбором	13
3 Технологический раздел	14
3.1 Средства реализации	14
3.2 Сведения о модулях программы	14
3.3 Реализация алгоритмов	14
3.4 Функциональные тесты	18
4 Исследовательский раздел	19
4.1 Демонстрация работы программы	19
4.2 Технические характеристики	20
4.3 Время выполнения реализаций алгоритмов	20
ЗАКЛЮЧЕНИЕ	27

ВВЕДЕНИЕ

Сортировка данных является фундаментальной задачей в области информатики и алгоритмов. Независимо от конкретной области применения, эффективные алгоритмы сортировки существенно влияют на производительность программных систем. От правильного выбора алгоритма зависит как время выполнения программы, так и затраты ресурсов компьютера [knut].

Алгоритмы сортировки находят применение в следующих сферах:

- базы данных;
- анализ данных и статистика;
- алгоритмы машинного обучения;
- криптография.

Цель данной лабораторной работы — рассмотреть алгоритмы сортировки. Для достижения поставленной цели необходимо выполнить следующие задачи:

- описать алгоритмы блочной, быстрой сортировок и сортировку выбором;
- разработать программное обеспечение, реализующее алгоритмы сортировок;
- выбрать инструменты для реализации и замера процессорного времени выполнения реализаций алгоритмов;
- проанализировать затраты реализаций алгоритмов по времени.

1 Аналитический раздел

Сортировкой называют перестановку объектов, при которой они располагаются в порядке возрастания или убывания [**knut**].

В данном разделе будут описаны три алгоритма сортировок: блочная, быстрая и выбором.

1.1 Алгоритм блочной сортировки

Идея заключается в разбиении входных данных на «блоки» одинакового размера, после чего данные в блоках сортируются и результаты сортировок объединяются. Отсортированная последовательность получается путём последовательного перечисления элементов каждого блока. Для деления данных на блоки, алгоритм предполагает, что значения распределены равномерно, и распределяет элементы по блокам равномерно. Например, предположим, что данные имеют значения в диапазоне от 1 до 100 и алгоритм использует 10 блоков. Алгоритм помещает элементы со значениями 1–10 в первый блок, со значениями 11–20 во второй, и т.д. Если элементы распределены равномерно, в каждый блок попадает примерно одинаковое число элементов. Если в списке N элементов, и алгоритм использует N блоков, в каждый блок попадает всего один или два элемента, поэтому возможно отсортировать элементы за конечное число шагов [**article_sorts**].

1.2 Алгоритм быстрой сортировки

Данный алгоритм можно разделить на следующие шаги [**quicksort**]:

- разбить массива относительно опорного элемента;
- рекурсивно отсортировать каждую часть массива:

1.3 Алгоритм сортировки выбором

Данный алгоритм можно разделить на следующие шаги [**selectionsort**]:

- взять первый элемент последовательности $A[i]$, здесь i – номер элемента, для первого i равен 1;
- найти минимальный (максимальный) элемент последовательности и запомнить его номер в переменную `key`;

- если номер первого элемента и номер найденного элемента не совпадают, т. е. если $key \neq 1$, тогда два этих элемента обменять значениями;
- увеличить i на 1 и продолжить сортировку оставшейся части массива, а именно с элемента с номером 2 по N , так как элемент $A[1]$ уже занимает свою позицию.

Вывод

В данном разделе были описаны три алгоритма сортировок: блочная, быстрая и выбором.

2 Конструкторский раздел

В этом разделе будет представлено описание используемых типов данных, а также схематические изображения алгоритмов сортировок: блочной, быстрой и выбором.

2.1 Требования к программному обеспечению

Программа должна поддерживать два режима работы: режим массового замера времени и режим сортировки введенного массива.

Режим массового замера времени должен обладать следующей функциональностью:

- генерировать массивы различного размера для проведения замеров;
- осуществлять массовый замер, используя сгенерированные данные;
- результаты массового замера должны быть представлены в виде таблицы и графика.

К режиму сортировки выдвигается следующий ряд требований:

- возможность работать с массивами разного размера, которые вводит пользователь;
- наличие интерфейса для выбора действий;
- на выходе программы массив, отсортированный тремя алгоритмами по возрастанию.

2.2 Описание используемых типов данных

При реализации алгоритмов будут использованы следующие структуры и типы данных:

- целое число представляет количество элементов в массиве;
- список целых чисел;

2.3 Разработка алгоритмов

На рисунке 2.1 приведена схема алгоритма блочной сортировки.

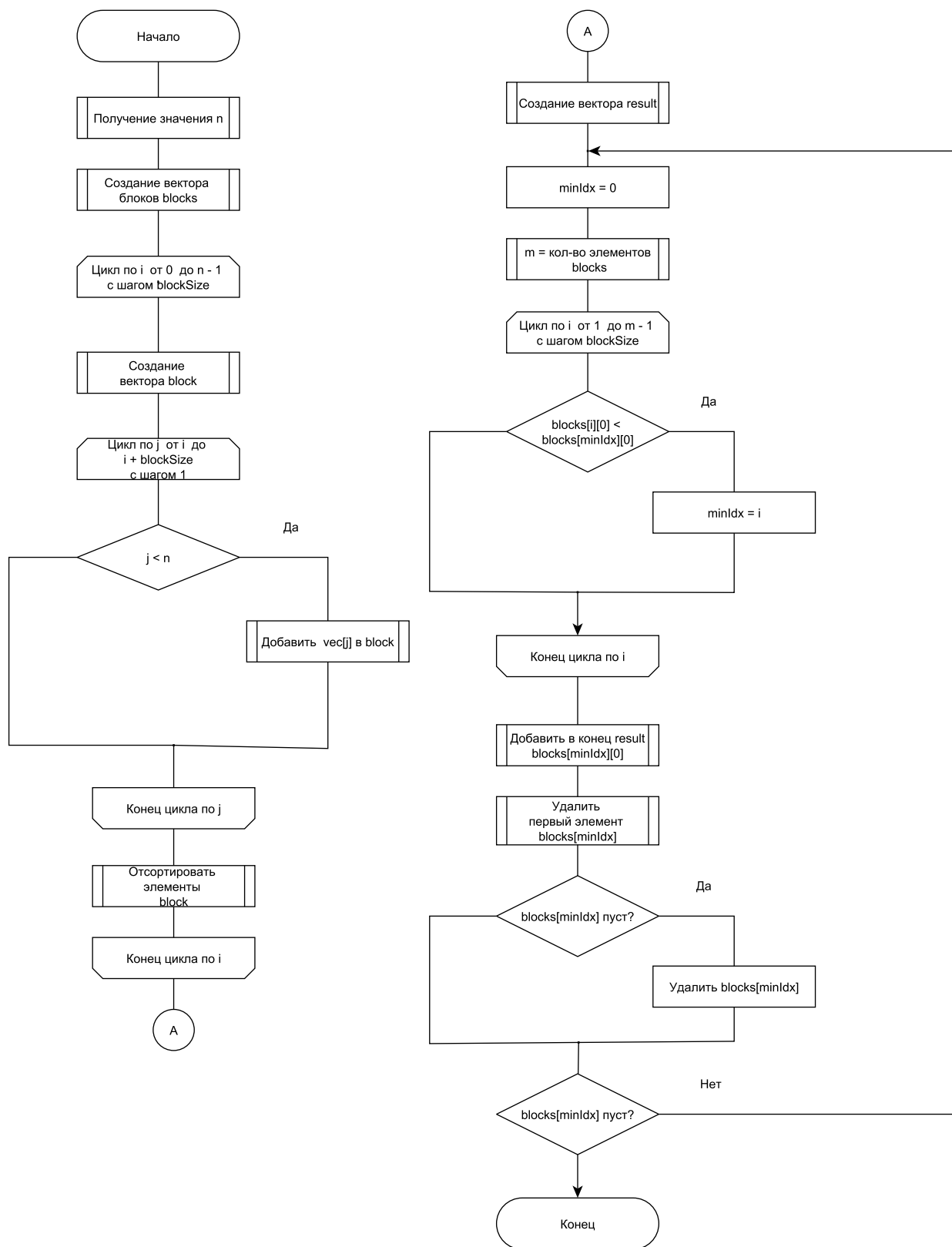


Рисунок 2.1 – Схема алгоритма блочной сортировки

На рисунке 2.2 приведена схема алгоритма быстрой сортировки.

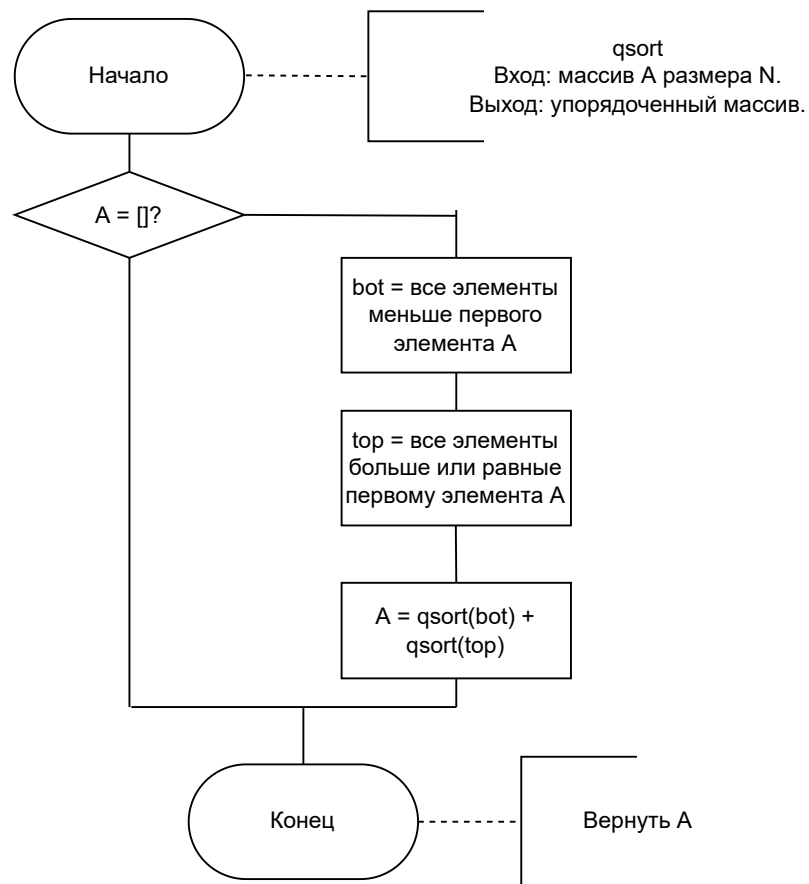


Рисунок 2.2 – Схема алгоритма быстрой сортировки

На рисунке 2.3 приведена схема алгоритма сортировки выбором.

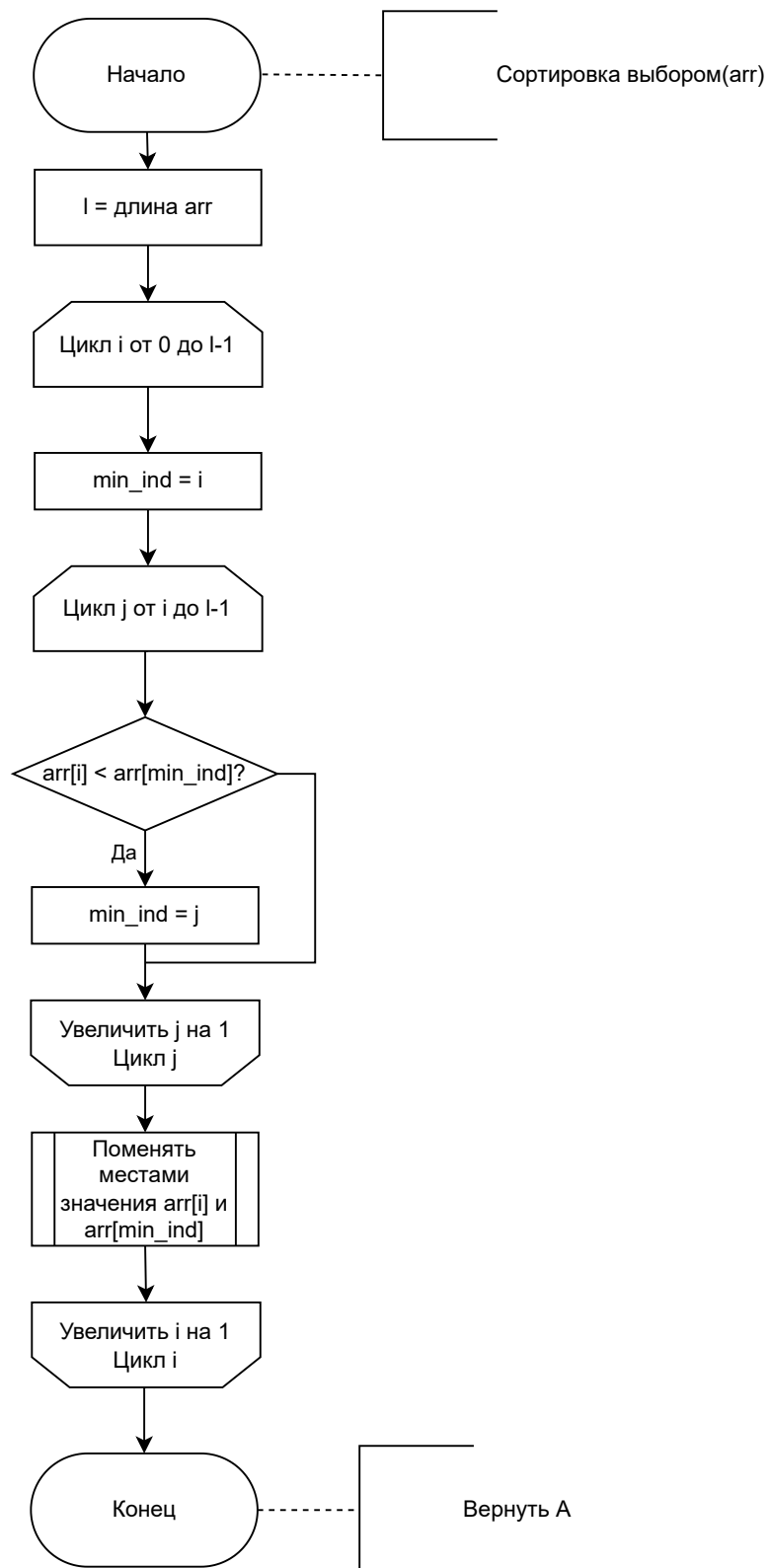


Рисунок 2.3 – Схема алгоритма сортировки выбором

2.4 Оценка трудоемкости алгоритмов

Модель для оценки трудоемкости алгоритмов состоит из шести пунктов:

- 1) $+, -, =, + =, - =, ==, ||, \&\&, <, >, <=, >=, <<, >>, []$ — считается, что эти операции обладают трудоемкостью в 1 единицу;
- 2) $*, /, * =, / =, \%$ — считается, что эти операции обладают трудоемкостью в 2 единицы;
- 3) трудоемкость условного перехода принимается за 0;
- 4) трудоемкость условного оператора рассчитывается по формуле (2.1),

$$f_{if} = f_{\text{условия}} + \begin{cases} \min(f_1, f_2), & \text{лучший случай} \\ \max(f_1, f_2), & \text{худший случай} \end{cases}, \quad (2.1)$$

где f_1 — трудоемкость блока, который вычисляется при выполнении условия, а f_2 — трудоемкость блока, который вычисляется при невыполнении условия;

- 5) трудоемкость цикла рассчитывается по формуле (2.2),

$$f_{for} = f_{\text{инициализация}} + f_{\text{сравнения}} + M_{\text{итераций}} \cdot (f_{\text{тело}} + f_{\text{инкремент}} + f_{\text{сравнения}}); \quad (2.2)$$

- 6) вызов подпрограмм и передача параметров принимается за 0.

2.4.1 Трудоемкость алгоритма блочной сортировки

В данной реализации размер блока обозначается как k , трудоемкость операции добавления и удаления элемента из вектора равна 2.

Лучший случай: массив отсортирован, элементы распределены равномерно (все блоки содержат одинаковое число элементов), расчет трудоемкости данного случая приведен в (2.3).

$$\begin{aligned}
f_{best} &= 1 + 1 + \frac{n}{k} \cdot (1 + 2 + f_{shaker} + 2 + 1 + 4 + \\
&\quad + k \cdot (3 + 1 + 4)) + 1 + 1 + \\
&+ \frac{n}{k} \cdot (1 + 4 + 1 + 1 + 5 + 1 + 4 + 1 + 1 + n \cdot (5)) = \\
&= 4 + \frac{29 \cdot n + n \cdot f_{shaker} + 5 \cdot n^2}{k} + 8 \cdot n = \\
&= 4 + 8 \cdot n + 29 \cdot \frac{n}{k} + n \cdot (14.5 + \frac{k}{2}) + \frac{5 \cdot n^2}{k} = \\
&= 4 + 8 \cdot n + 29 \cdot \frac{n}{k} + 14.5 \cdot n + \frac{n \cdot k}{2} + \frac{5 \cdot n^2}{k} = \\
&= \frac{5 \cdot n^2}{k} + 22.5 \cdot n + 29 \cdot \frac{n}{k} + \frac{n \cdot k}{2} + 4.
\end{aligned} \tag{2.3}$$

Худший случай: большое количество пустых блоков, массив отсортирован в обратном порядке (худший случай сортировки перемешиванием, которая используется в блочной сортировке), расчет трудоемкости приведен в выражении (2.4).

$$\begin{aligned}
f_{worst} &= 1 + 1 + \frac{n}{k} \cdot (1 + 2 + f_{shaker} + 2 + 1 + 4 + \\
&\quad + k \cdot (3 + 1 + 4)) + 1 + 1 + \\
&+ \frac{n}{k} \cdot (1 + 4 + 1 + 1 + 5 + 1 + 4 + 1 + 1 + k \cdot (6)) = \\
&= 4 + \frac{29 \cdot n + n \cdot f_{shaker} + 6 \cdot n^2}{k} + 8 \cdot n = \\
&= 4 + 8 \cdot n + 29 \cdot \frac{n}{k} + n \cdot (19.5 + \frac{k}{2}) + \frac{6 \cdot n^2}{k} = \\
&= 4 + 8 \cdot n + 29 \cdot \frac{n}{k} + 19.5 \cdot n + \frac{n \cdot k}{2} + \frac{6 \cdot n^2}{k} = \\
&= \frac{6 \cdot n^2}{k} + 27.5 \cdot n + 29 \cdot \frac{n}{k} + \frac{n \cdot k}{2} + 4.
\end{aligned} \tag{2.4}$$

2.4.2 Трудоемкость алгоритма быстрой сортировки

Чтобы вычислить трудоемкость алгоритма быстрой сортировки, нужно учесть следующее:

- трудоемкость условного оператора на проверку *pivot* в теле цикла вычисляется как 2.5

$$f_{if_pivot} = 1 + \begin{cases} 2 + 1, \\ 0 \end{cases} \quad (2.5)$$

- трудоемкость цикла, в теле которого условный оператор на проверку *pivot*, указана в формуле 2.7

$$f_{for_low_high} = 1 + M \cdot f_{if_pivot} \quad (2.6)$$

- трудоемкость условного оператора на проверку *low* в теле цикла вычисляется как 2.7

$$f_{if_low} = 1 + \begin{cases} 1, \\ 0 \end{cases} \quad (2.7)$$

- трудоемкость цикла, в теле оператор ветвления f_{if_low} и $f_{for_low_high}$, указана в формуле 2.8

$$f_{while_stack} = 1 + N \cdot (2 + f_{if_low} + 2 + f_{for_low_high} + 2 + 2) \quad (2.8)$$

- трудоемкость условного оператора на проверку длины массива вычисляется как 2.9

$$f_{if_len} = 1 + \begin{cases} 1, \\ 1 + f_{while_stack} \end{cases} \quad (2.9)$$

В итоге, трудоемкость быстрой сортировки вычисляется как 2.10:

$$\begin{aligned} f_{quick_sort} &= 1 + f_{if_len} = 1 + 1 + N \cdot (2 + f_{if_low} + 2 + f_{for_low_high}) = \\ &= 1 + 1 + N \cdot (2 + 2 + 2 + 1 + M \cdot f_{if_pivot} + 2 + 2) = \\ &= 1 + 1 + N \cdot (2 + 2 + 2 + 1 + M \cdot (3 + 2 + 2)) + 1 + 3 + 2 + 1 = \\ &= 2 + 7N + 7MN + 7N = 14N + 7MN \quad (2.10) \end{aligned}$$

2.4.3 Трудоемкость алгоритма сортировки выбором

Трудоемкость сортировки выбором в худшем случае $O(N^2)$

Трудоемкость сортировки выбором в лучшем случае $O(N^2)$

Вывод

На основе теоретических данных, полученных из аналитического раздела были построены схемы требуемых алгоритмов. Была введена модель оценки трудоемкости алгоритма, были рассчитаны трудоемкости алгоритмов в соответствии с этой моделью.

В результате теоретической оценки трудоемкостей алгоритмов выяснилось, что лучшей асимптотической оценкой во всех случаях обладает быстрая сортировка.

Худшей асимптотической оценкой во всех случаях обладает сортировка выбором $O(N^2)$.

3 Технологический раздел

В данном разделе будут приведены требования к программному обеспечению, средства реализации, листинг кода и функциональные тесты.

3.1 Средства реализации

Для реализации данной работы был выбран язык *Python* [python]. Данный выбор обусловлен следующим:

- язык поддерживает все структуры данных, которые выбраны в результате проектирования;
- язык позволяет реализовать все алгоритмы, выбранные в результате проектирования;
- язык позволяет замерять процессорное время с помощью модуля *time*.

Процессорное время было замерено с помощью функции *process_time()* из модуля *time* [python-time].

3.2 Сведения о модулях программы

Данная программа разбита на следующие модули:

- *main.py* — файл, содержащий функцию *main*;
- *functions.py* — файл, содержащий вспомогательные функции (ввод массива с клавиатуры, рандомно, с файла и т.д.)
- *sorts.py* — файл, содержащий код реализаций всех алгоритмов сортировок;
- *tests.py* — файл, в котором содержатся функции для замера и вывода времени выполнения реализаций алгоритмов.

3.3 Реализация алгоритмов

В листингах 3.1 – 3.3 приведены реализации блочной, быстрой сортировок и сортировки выбором.

Листинг 3.1 – Реализация блочной сортировки

```
1 def block_sort(arr):
2     block_size = 10
3     blocks = []
4
5     i = 0
6     while i < len(arr):
7         block = []
8         j = i
9         while j < i + block_size and j < len(arr):
10             block.append(arr[j])
11             j += 1
12         block.sort()
13         blocks.append(block)
14         i += block_size
15
16     arr_ind = 0
17     while blocks:
18         min_ind = 0
19         for i in range(1, len(blocks)):
20             if blocks[i][0] < blocks[min_ind][0]:
21                 min_ind = i
22
23         arr[arr_ind] = blocks[min_ind][0]
24         arr_ind += 1
25         blocks[min_ind].pop(0)
26
27         if not blocks[min_ind]:
28             blocks.pop(min_ind)
29
30     return arr
```

Листинг 3.2 – Реализация быстрой сортировки

```
1 import random
2
3 def quick_sort(arr):
4     less = []
5     equal = []
6     greater = []
7
8     if len(arr) > 1:
9         pivot = arr[random.randint(0, len(arr) - 1)]
10        for x in arr:
11            if x < pivot:
12                less.append(x)
13            elif x == pivot:
14                equal.append(x)
15            elif x > pivot:
16                greater.append(x)
17        return quick_sort(less) + equal + quick_sort(greater)
18
19 else:
20     return arr
```


Листинг 3.3 – Реализация сортировки выбором

```
1 def selection_sort(arr):  
2     size = len(arr)  
3  
4     for ind in range(size):  
5         min_ind = ind  
6  
7         for j in range(ind + 1, size):  
8             if arr[j] < arr[min_ind]:  
9                 min_ind = j  
10        arr[ind], arr[min_ind] = arr[min_ind], arr[ind]  
11  
12    return arr
```

3.4 Функциональные тесты

В таблице 3.1 приведены функциональные тесты для разработанных алгоритмов сортировок. Все тесты пройдены успешно.

Таблица 3.1 – Функциональные тесты

Массив	Ожидаемый результат	Фактический результат
[1, 2, 3, 4, 5]	[1, 2, 3, 4, 5]	[1, 2, 3, 4, 5]
[5, 4, 3, 2, 1]	[1, 2, 3, 4, 5]	[1, 2, 3, 4, 5]
[]	[]	[]
[1]	[1]	[1]
[4, 1, 2, 3]	[1, 2, 3, 4]	[1, 2, 3, 4]
[2, 1]	[1, 2]	[1, 2]
[31, 57, 24, -10, 59]	[-10, 24, 31, 57, 59]	[-10, 24, 31, 57, 59]

Вывод

Были разработаны и протестированы спроектированные алгоритмы сортировок: блочная, быстрая и выбором.

4 Исследовательский раздел

В данном разделе будут приведены: пример работы программы, постановка эксперимента и сравнительный анализ алгоритмов на основе полученных данных.

4.1 Демонстрация работы программы

На рисунке 4.1 представлена демонстрация работы разработанного программного обеспечения, а именно показаны результаты сортировки массива $[3, 5, -3, 1, 2, 4, -6, 2]$.

```
МЕНЮ:
1. Блочная сортировка
2. Быстрая сортировка
3. Сортировка выбором
4. Измерить время
0. Выход

Выберите пункт: 1
Введите элементы массива (в одну строку через пробел):
3 5 -3 1 2 4 -6 2

Результат: [-6, -3, 1, 2, 2, 3, 4, 5]

МЕНЮ:
1. Блочная сортировка
2. Быстрая сортировка
3. Сортировка выбором
4. Измерить время
0. Выход

Выберите пункт: 2
Введите элементы массива (в одну строку через пробел):
3 5 -3 1 2 4 -6 2

Результат: [-6, -3, 1, 2, 2, 3, 4, 5]

МЕНЮ:
1. Блочная сортировка
2. Быстрая сортировка
3. Сортировка выбором
4. Измерить время
0. Выход

Выберите пункт: 3
Введите элементы массива (в одну строку через пробел):
3 5 -3 1 2 4 -6 2

Результат: [-6, -3, 1, 2, 2, 3, 4, 5]
```

Рисунок 4.1 – Демонстрация работы программы при сортировке массива

4.2 Технические характеристики

Технические характеристики компьютера, на котором проводился замерный эксперимент:

- процессор Intel Core i5-10400F (6 ядер) [**intel**];
- 16 Гб оперативная память DDR4;
- операционная система Windows 10 Pro [**windows**].

Во время проведения исследования компьютер был нагружен только системными приложениями и целевой программой.

4.3 Время выполнения реализаций алгоритмов

Результаты замеров времени выполнения реализаций алгоритмов сортировок приведены в таблицах 4.1 – 4.3. Замеры времени проводились на массивах одного размера и усреднялись для каждого набора одинаковых экспериментов.

В таблицах 4.1 – 4.3 используются следующие обозначения:

- Блочная — реализация алгоритма блочной сортировки;
- Быстрая — реализация алгоритма быстрой сортировки;
- Выбором — реализация алгоритма сортировки выбором.

Таблица 4.1 – Время работы реализации алгоритмов на неотсортированных массивах (в мс)

Размер массива	Блочная	Быстрая	Выбором
100	0.078125	0.125	0.203125
200	0.203125	0.265625	0.90625
300	0.40625	0.5	1.78125
400	0.765625	0.546875	3.484375
500	1.078125	0.734375	5.625
600	1.390625	0.890625	7.8125
700	1.8125	1.15625	10.609375
800	2.234375	1.25	14.25
900	2.984375	1.296875	18.0625
1000	3.734375	1.5	22.515625

Таблица 4.2 – Время работы реализации алгоритмов на отсортированных в обратном порядке массивах (в мс)

Размер массива	Блочная	Быстрая	Выбором
100	0.078125	0.125	0.203125
200	0.203125	0.25	0.765625
300	0.40625	0.390625	1.9375
400	0.640625	0.546875	3.3475
500	0.96875	0.6875	5.375
600	1.328125	0.875	7.609375
700	1.75	1.046875	10.640625
800	2.609375	1.140625	14.046875
900	2.9375	1.25	17.828125
1000	3.71875	1.4375	22.265625

Таблица 4.3 – Время работы реализации
алгоритмов на отсортированных массивах (в мс)

Размер массива	Блочная	Быстрая	Выбором
100	0.0625	0.109375	0.203125
200	0.203125	0.3125	0.828125
300	0.390625	0.4375	1.859375
400	0.640625	0.578125	3.453125
500	1	0.6875	5.3125
600	1.3125	0.859375	7.90625
700	1.890625	1.03125	10.46875
800	2.28125	1.125	13.8125
900	2.9375	1.3125	18.109375
1000	3.4375	1.4375	22.421875

На рисунках 4.2 – 4.4 изображены графики зависимостей времени выполнения реализаций сортировок от размеров массивов.

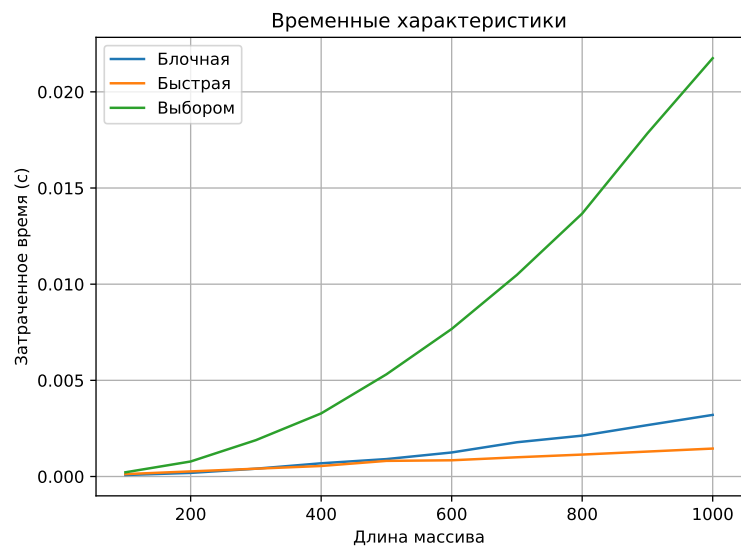


Рисунок 4.2 – Сравнение реализаций алгоритмов по времени выполнения на неотсортированных массивах

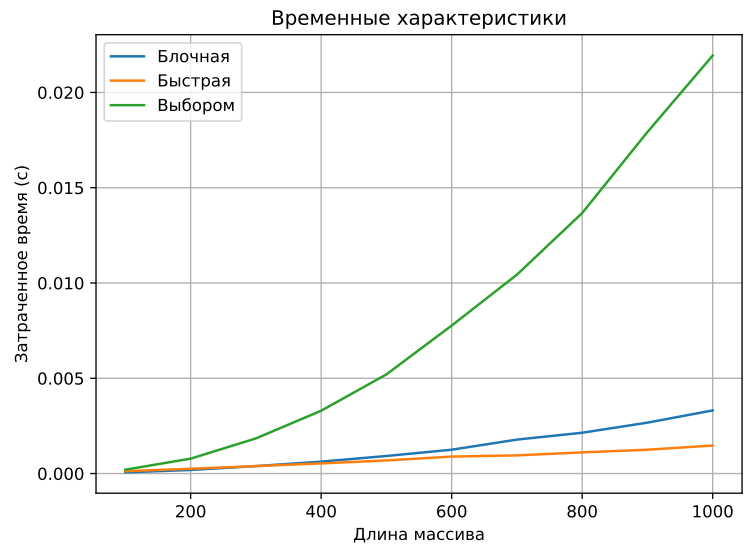


Рисунок 4.3 – Сравнение реализаций алгоритмов по времени выполнения на отсортированных в обратном порядке массивах

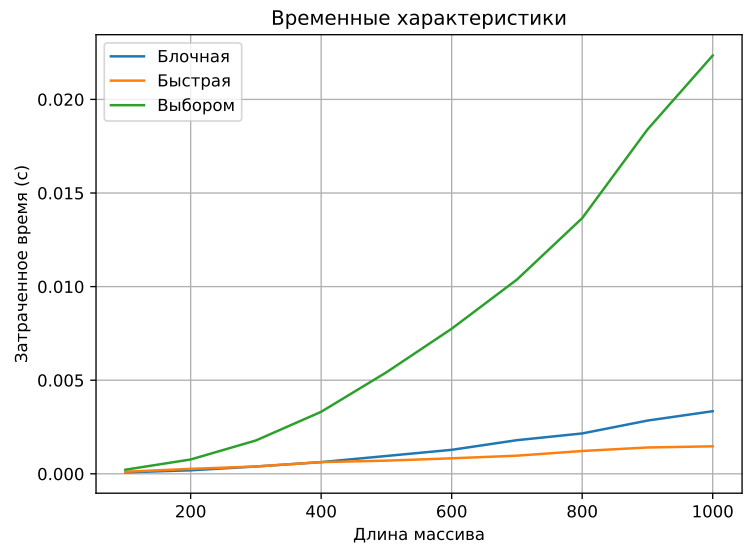


Рисунок 4.4 – Сравнение реализаций алгоритмов по времени выполнения на отсортированных массивах

Вывод

В результате замеров времени выполнения реализаций различных алгоритмов было выявлено, что для массивов длины 1000, отсортированных в обратном порядке, реализация алгоритма быстрой сортировки по времени оказалась в 2.59 раз лучше, чем реализация блочной сортировки, и в 15.49 раз лучше реализации сортировки выбором. В свою очередь, реализация блочной сортировки оказалась лучше в 5.99 раз по времени выполнения, чем реализация сортировки выбором.

Для отсортированных массивов длиной 1000 реализация быстрой сортировки оказалась лучше по времени в 2.39 раз, чем реализация блочной сортировки, и в 15.6 раз лучше, чем реализация сортировки выбором. В свою очередь, реализация блочной сортировки оказалась лучше в 6.52 раз по времени выполнения, чем реализация сортировки выбором.

Для случайно упорядоченных массивов длиной 1000 реализация быстрой сортировки оказалась лучше по времени в 2.49 раз, чем реализация блочной сортировки, и в 15.01 раза лучше, чем реализация сортировки выбором. В свою очередь, реализация блочной сортировки на случайно упорядоченных массивах оказалась лучше в 6.03 раза по времени выполнения, чем реализация сортировки выбором.

Стоит заметить, что для массивов длиной менее 300, реализация блочной сортировки была лучше или такой же по времени выполнения по сравнению с быстрой сортировкой, но на массивах длины более 300 быстрая сортировка становилась лучше по времени выполнению.

ЗАКЛЮЧЕНИЕ

Цель данной лабораторной работы была достигнута, а именно были рассмотрены алгоритмы сортировок.

Для достижения поставленной цели были выполнены следующие задачи.

- описаны алгоритмы блочной, быстрой сортировок и сортировки выбором;
- разработано программное обеспечение, реализующее алгоритмы сортировок;
- выбраны инструменты для реализации алгоритмов и замера процессорного времени их выполнения;
- проведен анализ затрат реализаций алгоритмов по времени.

В результате исследования реализаций различных алгоритмов было получено, что для массивов длины 1000, отсортированных в обратном порядке, реализация алгоритма быстрой сортировки по времени оказалась в 2.59 раз лучше, чем реализация блочной сортировки, и в 15.49 раз лучше реализации сортировки выбором. В свою очередь, реализация блочной сортировки оказалась лучше в 5.99 раз по времени выполнения, чем реализация сортировки выбором.

Для отсортированных массивов длиной 1000 реализация быстрой сортировки оказалась лучше по времени в 2.39 раз, чем реализация блочной сортировки, и в 15.6 раз лучше, чем реализация сортировки выбором. В свою очередь, реализация блочной сортировки оказалась лучше в 6.52 раз по времени выполнения, чем реализация сортировки выбором.

Для случайно упорядоченных массивов длиной 1000 реализация быстрой сортировки оказалась лучше по времени в 2.49 раз, чем реализация блочной сортировки, и в 15.01 раза лучше, чем реализация сортировки выбором. В свою очередь, реализация блочной сортировки на случайно упорядоченных массивах оказалась лучше в 6.03 раза по времени выполнения, чем реализация сортировки выбором.

Для массивов длиной менее 300, реализация блочной сортировки была лучше или такой же по времени выполнения по сравнению с быстрой сортировкой, но на массивах длины более 300 быстрая сортировка становилась лучше по времени выполнению.