**Московский государственный технический**

**университет им. Н.Э. Баумана**

Факультет «Информатика и системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Отчет по домашнему заданию №1

**«Анализ данных»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Выполнил:** | |  | **Принял:** | |
| ФИО: | \_Цыпышев Т. А.\_\_\_\_\_ |  | ФИО: | \_Семёнов Д. В.\_\_\_\_\_\_ |
| Группа: | \_ИУ5-11Б\_\_\_\_\_\_\_\_\_\_ |  | Должность: | \_Преподаватель\_\_\_\_\_ |
| Дата: | \_20.01.2023\_\_\_\_\_\_\_\_ |  | Дата: | \_20.01.2023\_\_\_\_\_\_\_\_\_ |
| Подпись: | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  | Подпись: | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Москва, 2022 г.

**Описание алгоритма**

**Алгоритм:**

Файл считывается в buffer, далее вызывается функция EditItem, в которой buffer побитно сдвигается универсально для каждого type и вызывается функция EditSet. В ней программа в зависимости от наличия buffer в set либо добавляет его, либо увеличивает кол-во (сколько раз buffer встречался в файле). Далее программа сортирует полученный set по 2 элементу с помощью функции SortingSet и выводит отсортированный set на экран с помощью функции Output.

**Дополнительно:**

В программе реализован метод без считывания файл в массив, за счёт этого уменьшается нагрузка на компьютер.

**Описание функций**

1. **int main()** – здесь происходит вся магия, отсюда вызываются все функции, весь алгоритм находиться тут
2. **void EditItem(const int type, const int byte, int \*buffer, uint64\_t \*len\_buffer, int set[][2], int \*len\_set)**

**Описание:**

Эта функция универсально для каждого типа (по 8, 12, 16, 20 бит) преобразовывает buffer (побитово сдвигает в зависимости от типа) и вызывает функцию **EditSet**.

**Переменные:**

* **const int type** – тип (длина по которой необходимо делить buffer): 8, 12, 16, 20 бит
* **const int byte** – значение байта, который мы считали из файла и должны прибавить к buffer
* **int \*buffer** – переменная необходима для преобразования (выше упоминалась как buffer)
* **uint64\_t \*len\_buffer** – длина буфера
* **int set[][2]** – двумерный массив (передаётся в EditSet)
* **int \*len\_set** – длина set

1. **void EditSet(int buffer, uint64\_t len\_buffer, int n, int set[][2], int \*len\_set)**

**Описание:**

Эта функция добавляет или изменяет элементы в set. В зависимости от наличия buffer в set либо добавляет его, либо увеличивает кол-во (сколько раз buffer встречался в файле).

**Переменные:**

* **int buffer** – переменная, которую нужно добавить в set
* **uint64\_t len\_buffer** – длина буфера
* **int n** – равен type (используется для сдвига buffer)
* **int set[][2]** – двумерный массив, где set[][0] – buffer set[][0] – count
* **int \*len\_set** – длина set

1. **void SortingSet(int set[][2], int len\_set)**

**Описание:**

Эта функция сортирует set по 2 элементу (количеству) методом пузырька (в дальнейшем можно заменить на что-то более быстрое)

**Переменные:**

* **int set[][2]** – двумерный массив, который должен быть отсортирован
* **int len\_set** – длина set

1. **void Output(int set[][2], int size, int count)**

**Описание:**

Эта функция выводит set в консоль

**Переменные:**

* **int set[][2]** – двумерный массив, который необходимо вывести
* **int size** – длина set
* **int count** – равен type (используется для точного вывода set[][0] в двоичном формате)

**Описание переменных**

**FILE \*file** – поток файла для чтения

**uint64\_t text\_size** – длина файла

**int set8[][2], set12[][2], set16[][2], set20[][2]** – двумерные массивы со статистикой (для 8, 12, 16, 20 бит)

**int len\_set8, len\_set12, len\_set16, len\_set20** – индекс первого свободного элемента в set (для 8, 12, 16, 20 бит)

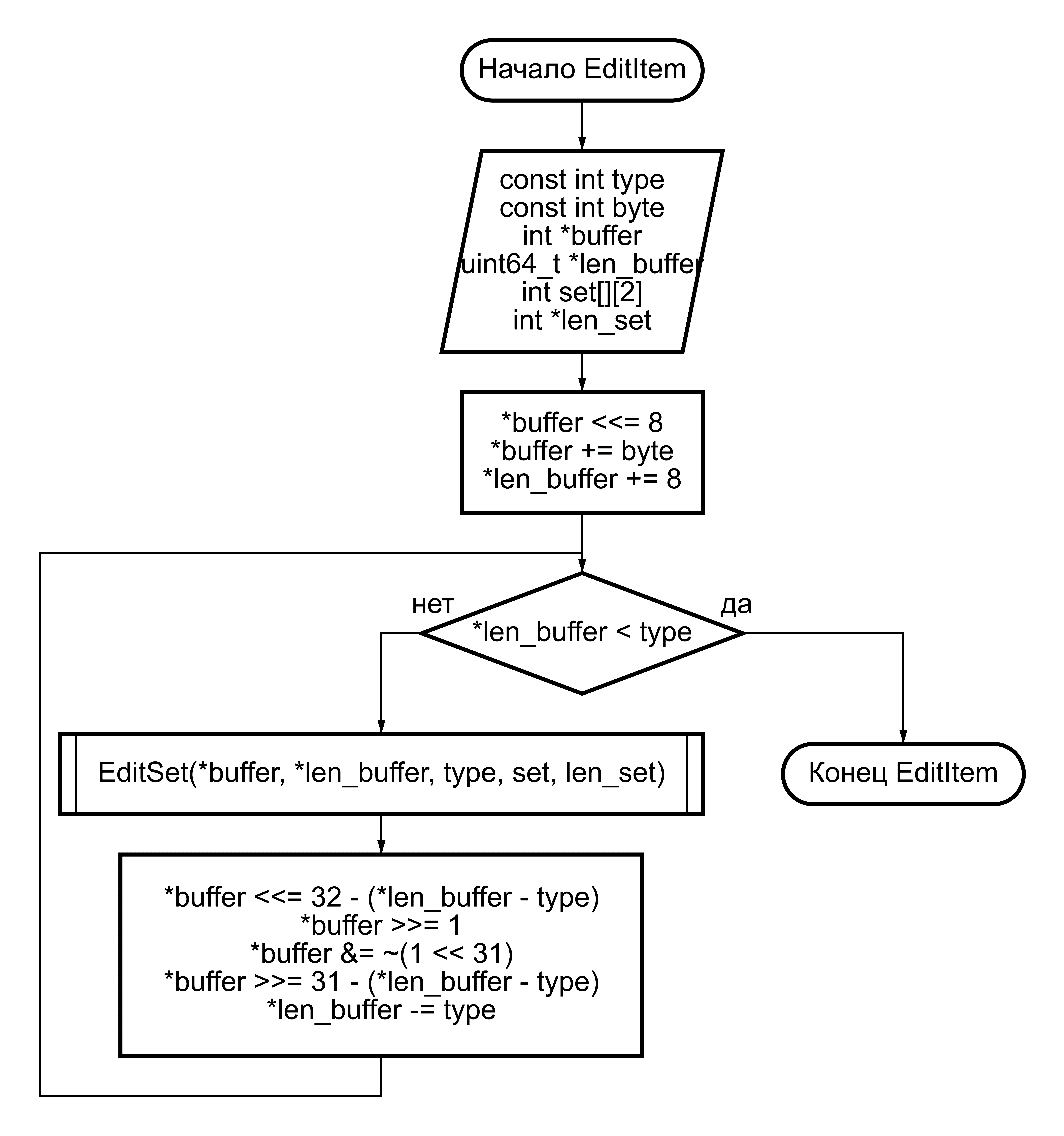
**int byte** – байт, который мы считываем из файла

**int buffer8 = 0, buffer12 = 0, buffer16 = 0, buffer20 = 0** – буфер, который в итоге добавляем в set (для 8, 12, 16, 20 бит)

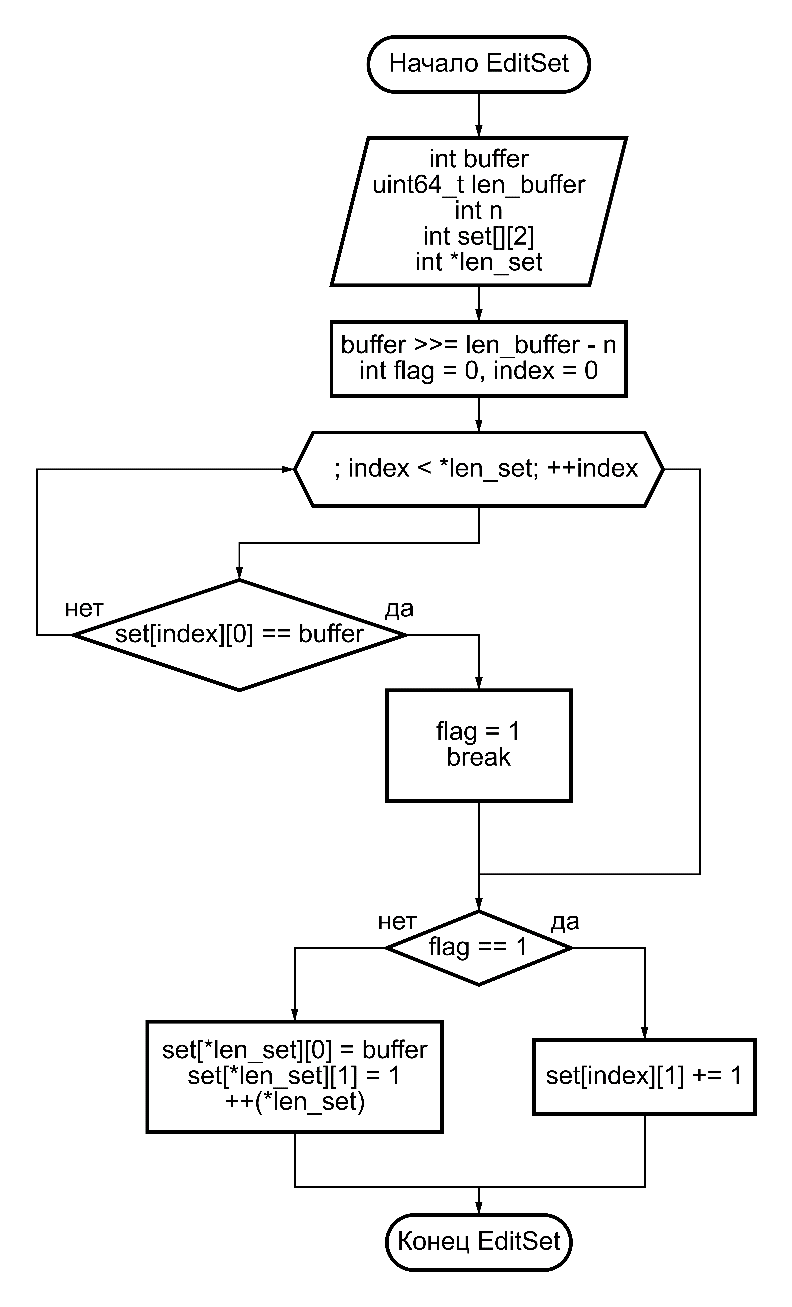
**uint64\_t len\_buffer8, len\_buffer12, len\_buffer16, len\_buffer20** – длина buffer

**Схема алгоритма**

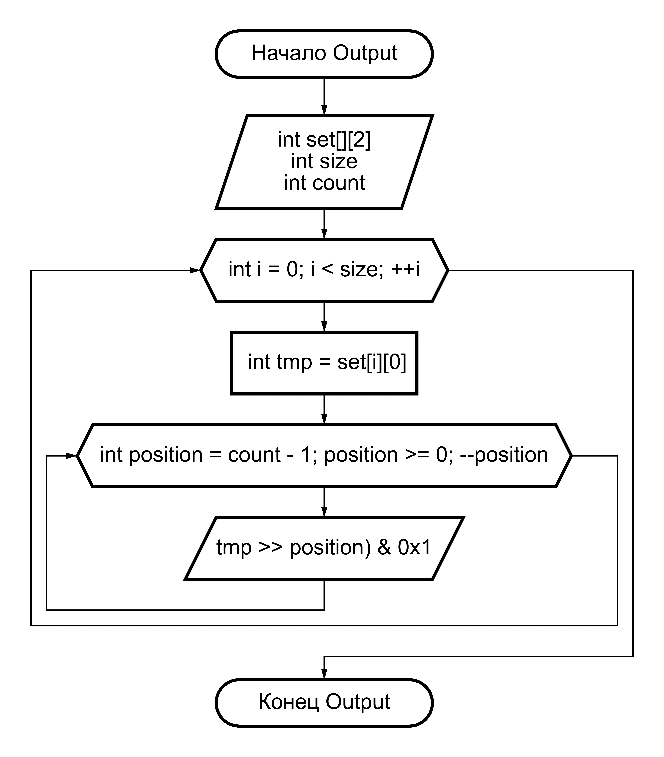
**EditItem**

****

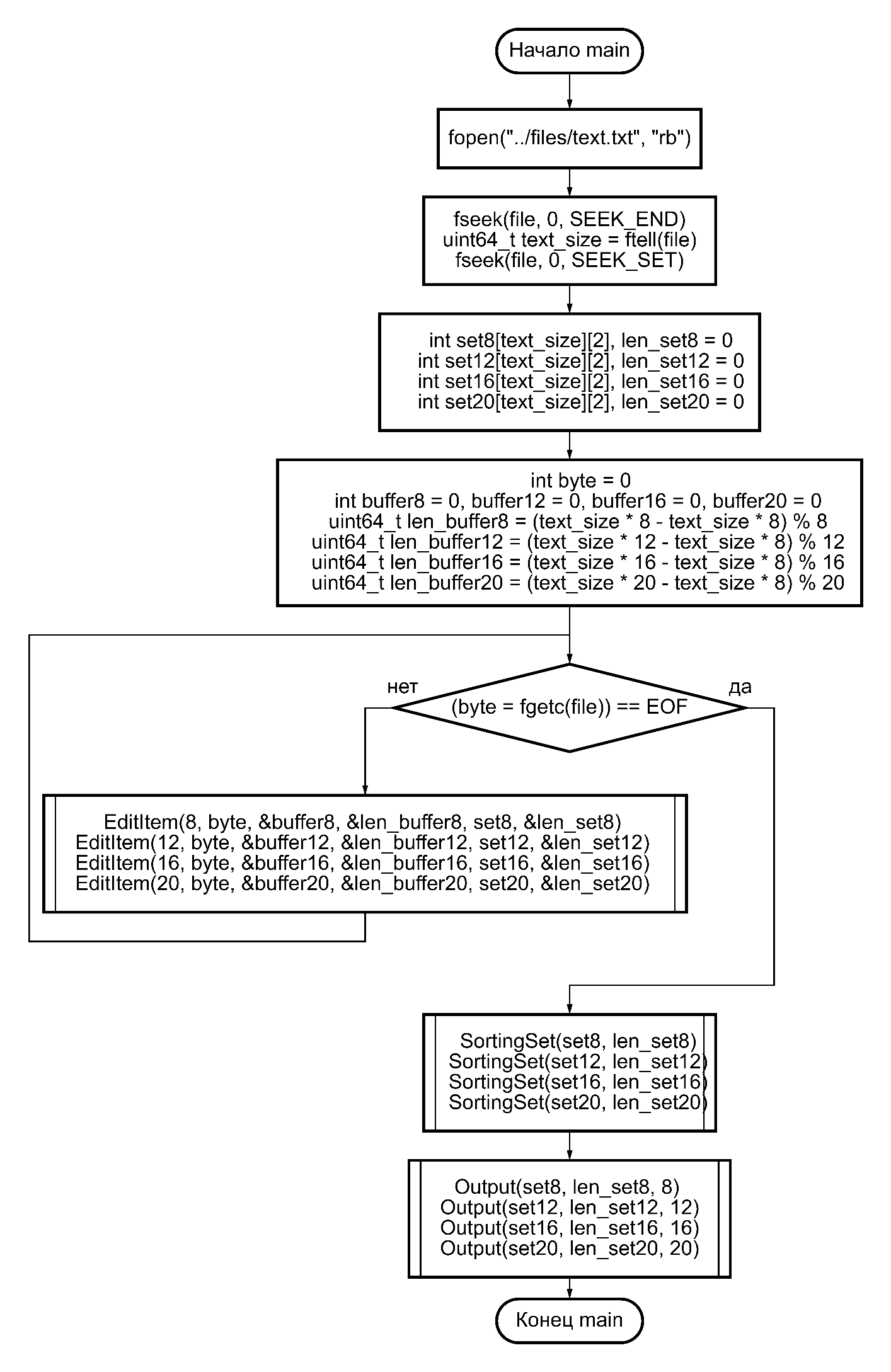
**EditSet**

****

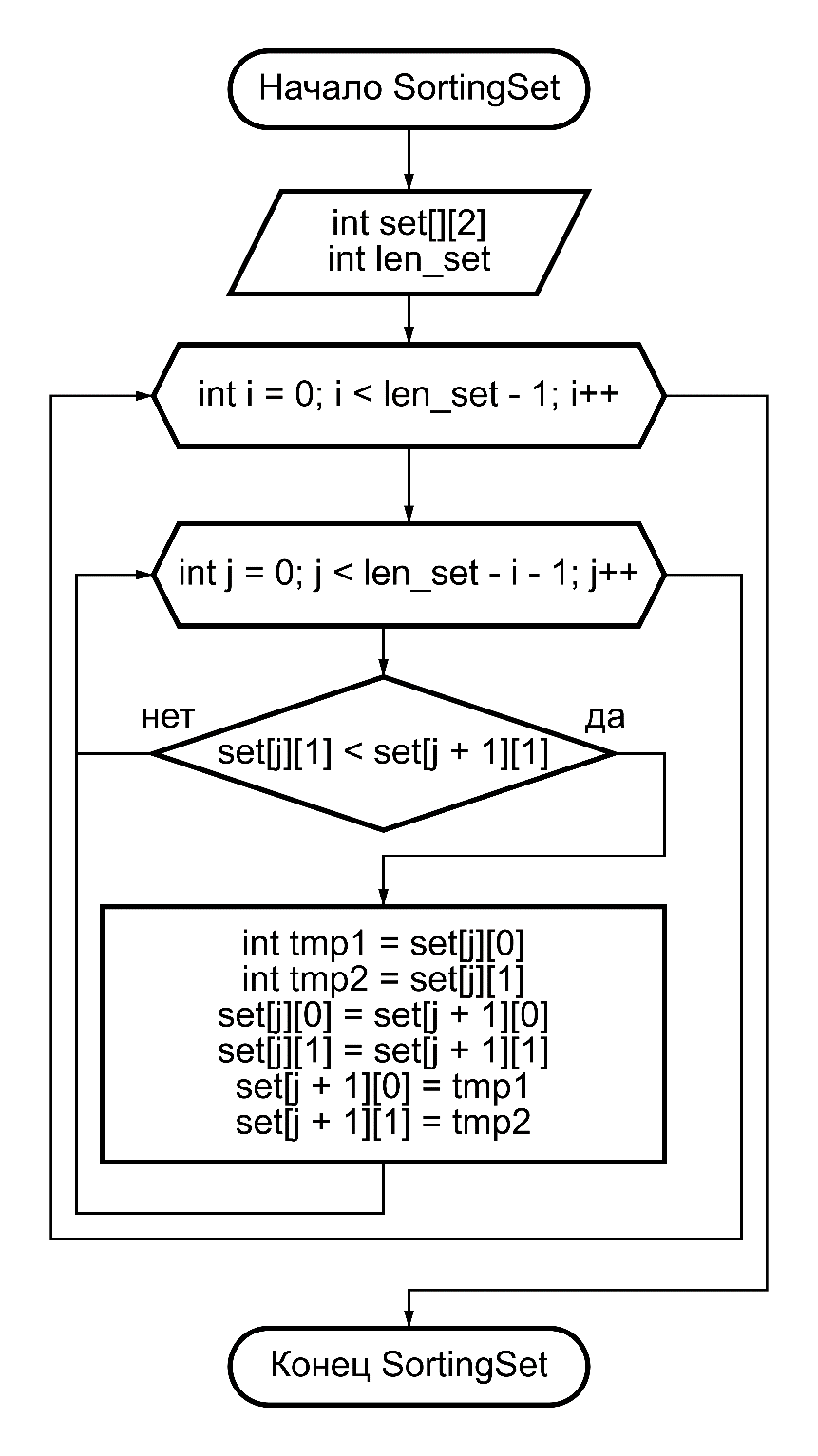
**Output**

****

**Main**

****

**SortingSet**

****