

МИНОБРНАУКИ РОССИИ

*Федеральное государственное бюджетное образовательное учреждение высшего образования*

***«МИРЭА – Российский технологический университет»***

РТУ МИРЭА

Отчет по выполнению практического задания № 4

**Тема:**

«Алгоритмы внешних сортировок»

Дисциплина: «Структуры и алгоритмы обработки данных»

Выполнил студент: Боргачев Т.М.

Группа: ИНБО-10-23

Москва – 2024

# СОДЕРЖАНИЕ

[1 ФОРМУЛИРОВКА ЗАДАЧИ 3](#_Toc161744244)

[1.1 Задание 1 3](#_Toc161744245)

[1.3 Задание 2 3](#_Toc161744246)

[2 ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ 3](#_Toc161744247)

[2.1 Алгоритм сортировки прямого слияния для файлов 4](#_Toc161744248)

[2.1.1 Реализация Алгоритма в виде функции 4](#_Toc161744249)

[2.2 Алгоритм естественного слияния 11](#_Toc161744250)

[2.2.1 Задание 2 13](#_Toc161744251)

[3 ВЫВОДЫ 20](#_Toc161744252)

[4 ИНФОРМАЦИОННЫЕ ИСТОЧНИКИ 20](#_Toc161744253)

Цель: освоить приёмы сортировки данных из файлов.

# 1 ФОРМУЛИРОВКА ЗАДАЧИ

## Задание 1

Разработать программу и применить алгоритм внешней сортировки прямого слияния к сортировке файла данных индивидуального варианта 5 по значению ключевого поля «Издательство».

1) Реализовать функцию сортировки (возможно, с вспомогательными функциями) и основную подпрограмму main.

2) Отладить программу, протестировать на примере.

3) Предварительно подготовить файл данных в соответствии с вариантом (не менее 32 записей).

4) Адаптировать программу для сортировки файла с записями, протестировать на подготовленном ранее файле.

5) Определить практическую сложность алгоритма для файлов с увеличивающимся количеством записей (8, 16, 32). Сформировать таблицу результатов, указав количество записей и время сортировки.

## 1.3 Задание 2

Разработать программу и применить алгоритм сортировки естественного слияния к сортировке файла с данными варианта (файл уже должен быть подготовлен в задании 1).

1) Реализовать функцию сортировки (возможно, с вспомогательными функциями) и основную подпрограмму main.

2) Отладить программу, протестировать на примере.

3) Адаптировать программу для сортировки файла с записями, протестировать на подготовленном ранее файле.

4) Сформировать таблицу результатов, указав количество записей и время сортировки.

# ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ

## Алгоритм сортировки прямого слияния для файлов

Фаза разделения:

1. Открыть файл А как входной.

2. Открыть файлы В и С как выходные (для записи).

3. Считываемые из А записи попеременно записываем в файлы В и С.

4. Закрываем файлы А, В, С.

Фаза слияния:

1. Открыть файл А как выходной (для записи).

2. Открываем файлы В и С как входные (для чтения).

3. Установить размер порции сливаемых данных: 1, 2, 4, 8 и т.д. для этого и следующих этапов.

4. Для каждой порции считываются по одной записи из файлов В и С. 5. Меньшая запись записывается в файл А, и считывается очередная запись из того файла, запись которого была переписана в файл А.

6. Пункты 4 и 5 повторяются до тех пор, пока записи очередной порции одного из файлов не будут исчерпаны.

7. Оставшиеся записи из порции другого файла переписываются в файл А.

8. Пункты с 4 по 7 повторяются до тех пор, пока не будет достигнут конец одного из файлов В и С. Тогда оставшиеся записи из другого файла переписываются в файл А.

9. Закрываются файлы А В С. Сортировка завершается тогда, когда длина порции достигнет n.

### 2.1.1 Реализация Алгоритма в виде функции

Реализация алгоритма в виде блок-схем представлена на рис. 1, 2 и 3.

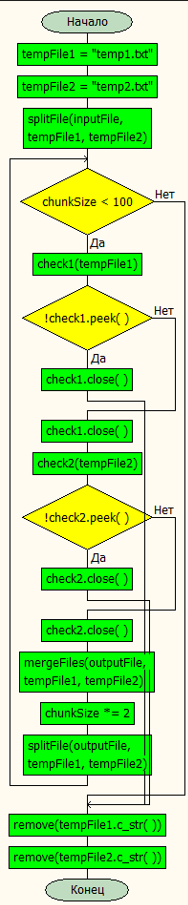
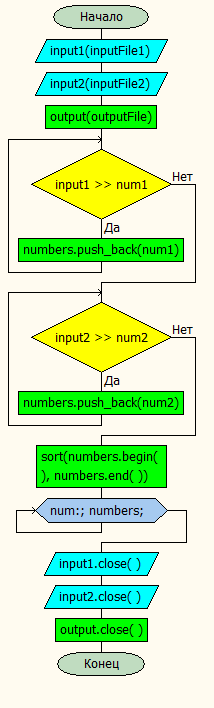


Рисунок 1 – Блок схема функции сортировки вставками для файлов



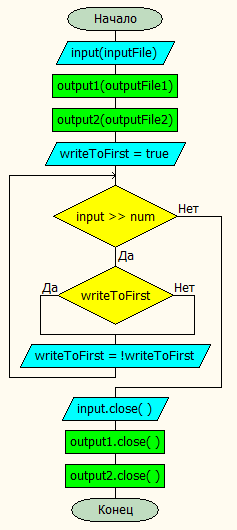
Рисунок 2 – Блок схема функции слияния данных из двух файлов

Рисунок 3 – Блок-схема функции разделения данных из файла на два

Код на языке C++, реализующий алгоритм слияния в виде функций представлен на рис. 4, 5 и 6.

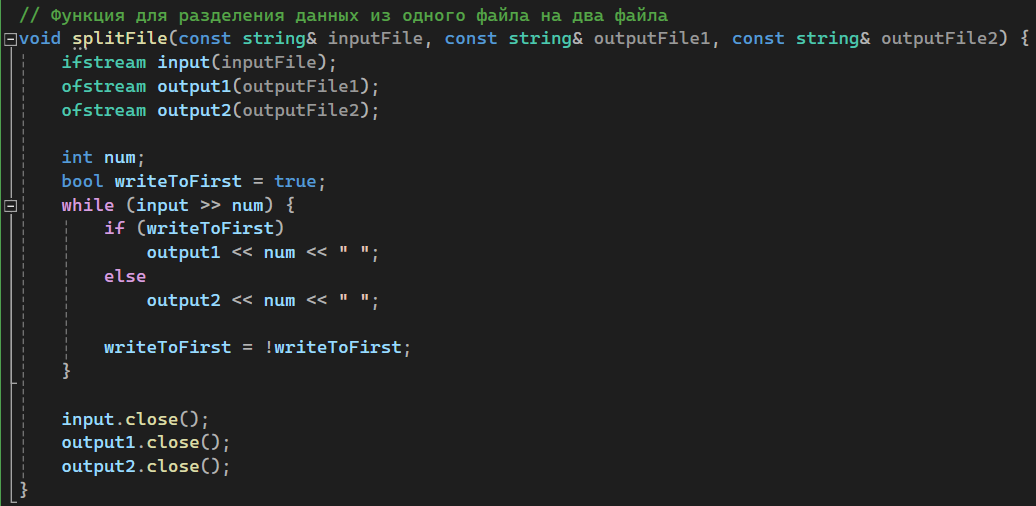


Рисунок 4 - Функция для разделения данных из одного файла на два файла

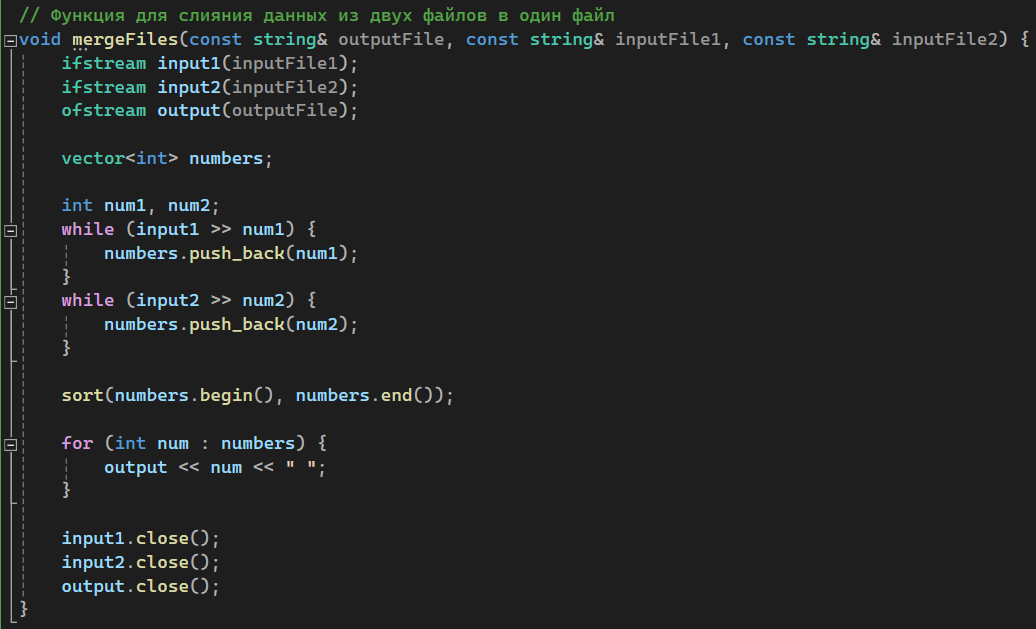


Рисунок 5 - Функция для слияния данных из двух файлов в один файл

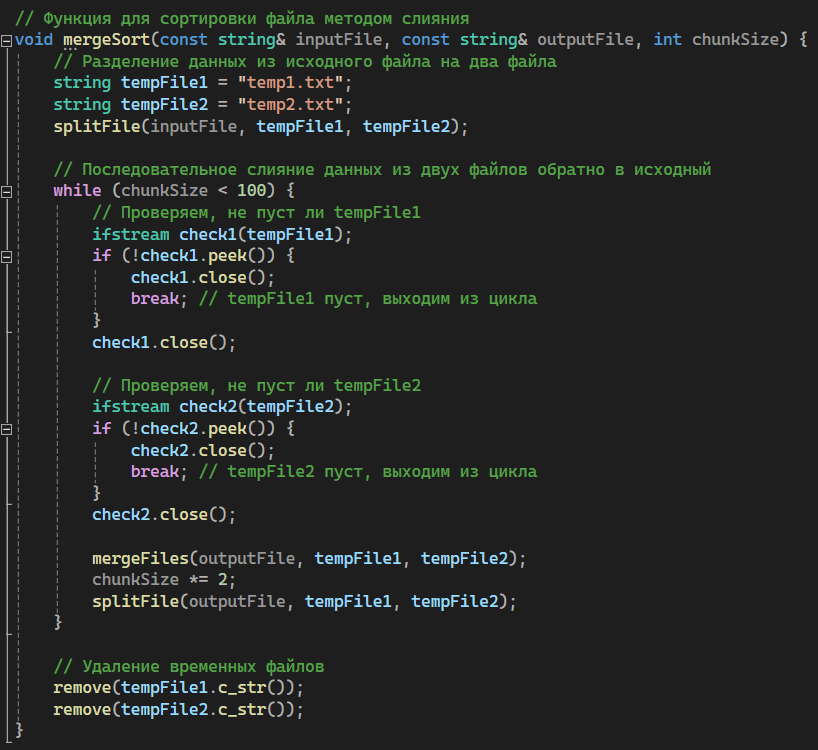


Рисунок 6 - Функция для сортировки файла методом слияния

Результаты тестирования алгоритма на примере (8 2 13 4 15 6 9 11 3 7 5 10 1 12 14) представлены на рис. 7.

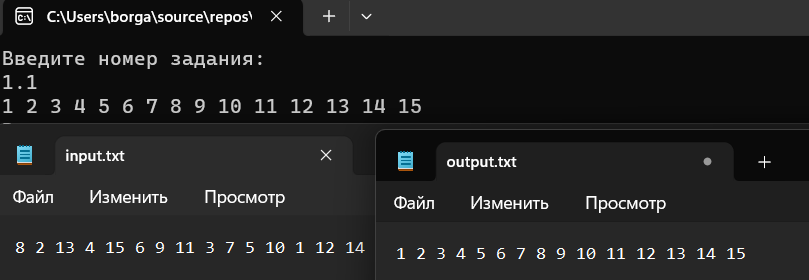


Рисунок 7 – Тестирования алгоритма на примере

Теперь необходимо модернизировать программу для работы с индивидуальным вариантом. Файл с данными представлен на рис. 8.

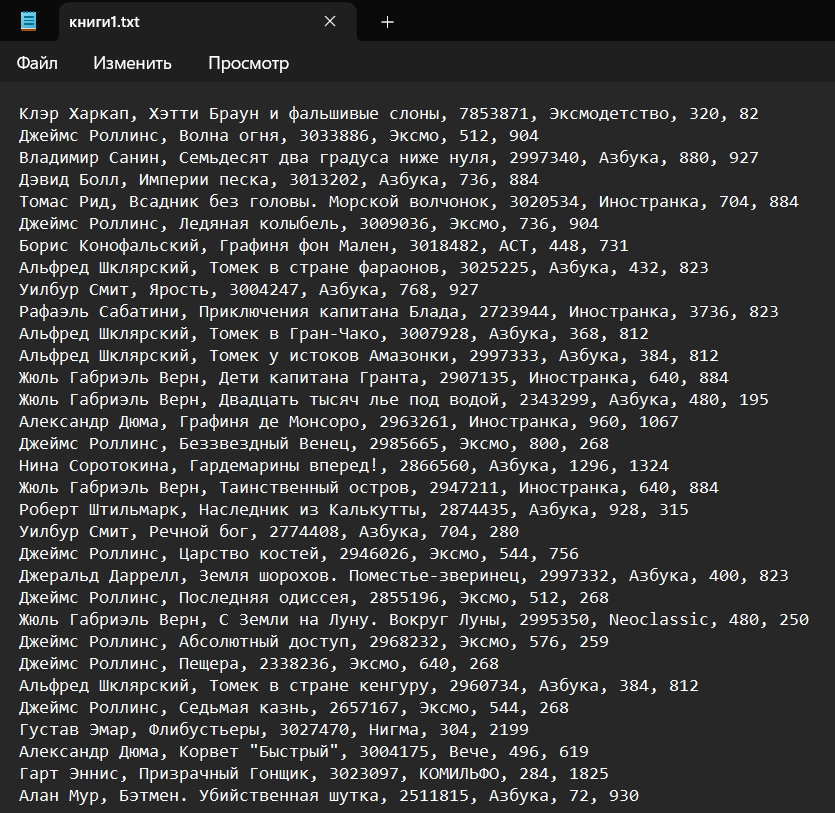


Рисунок 8 – Тестовые данные

Измененные и добавленные функции представлены на рис. 9.

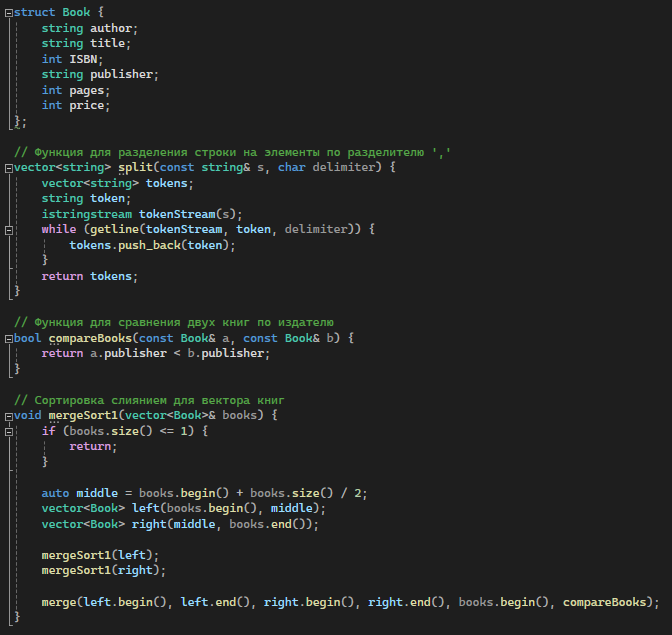


Рисунок 9 – Сортировка слиянием книг

Результаты тестирования алгоритма представлены на рис. 10.

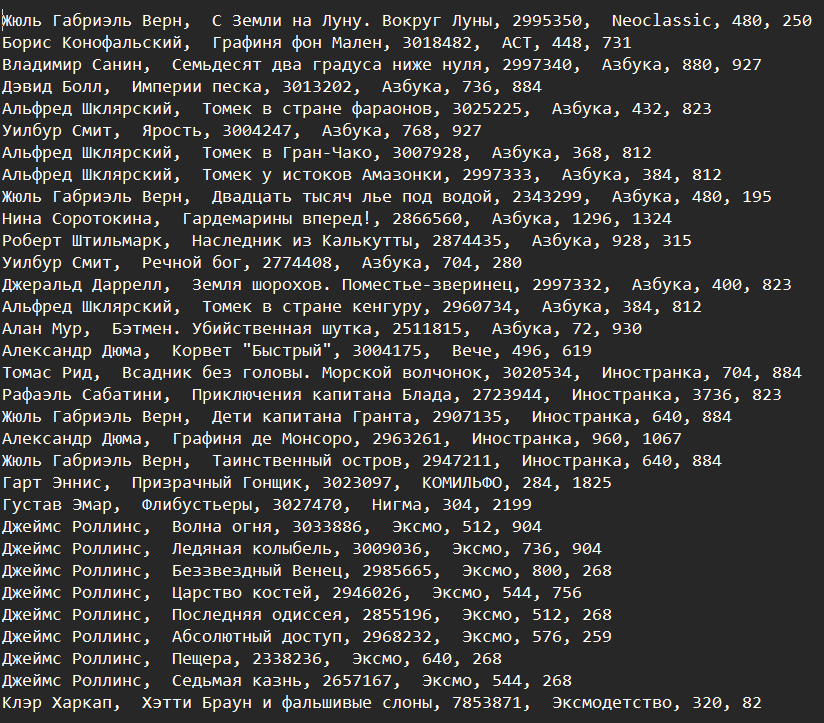


Рисунок 10 – Тестирование алгоритма

Практическая сложность алгоритма для различных количеств записей представлена в табл. 1.

Таблица 1 – Практическая сложность алгоритма прямого слияния

|  |  |
| --- | --- |
| Количество записей | Время работы алгоритма, мс |
| 8 | 11 |
| 16 | 16 |
| 32 | 24 |

## 2.2 Алгоритм естественного слияния

Сортировка естественного слияния, рассматривает две сливаемые подпоследовательности, как упорядоченные. Упорядоченные подпоследовательности принято называть сериями. Пусть исходный файл разделен на два файла, каждый из которых содержит по n – серий (один может содержать n-1 серию). Тогда при слиянии этих файлов будет получен файл из n серий. При каждом проходе число серий уменьшается вдвое, и общее число пересылок в худшем случае равно n log2n, а в среднем меньше. Процесс сортировки заканчивается, если при очередном проходе в файл будет перелита только одна серия.

Для усовершенствования этой сортировки был предложен вариант предварительного разделения данных в файле на серии одной длины, загрузки каждой серии в оперативную память, сортировки этой серии, например, алгоритмом быстрой сортировки, и запись этих серий в исходный файл. Чем длиннее серию возможно выгрузить в память, отсортировать и вернуть в файл, тем эффективнее будет алгоритм самой сортировки. Такое решение предлагается вам исследовать и разобраться в реализации. Рассмотрим алгоритм и его фазы. Он так же является двухфазным.

1. Определить размер свободной оперативной памяти для выгрузки в нее серии из файла. В программе создаем массив для хранения серии buf.

2. Открыть исходный файл А, подлежащий сортировке.

3. Открыть два файла для записи В и С.

4. Считать последовательность данных в количестве достаточном для размещения в массиве buf. Отсортировать в массиве методом внутренней сортировки и записать в файл В.

5. Считать следующую последовательность данных в количестве достаточном для размещения в массиве buf. Отсортировать в массиве методом внутренней сортировки и записать в файл С.

6. Пункты 4 и 5 выполнять, пока все данные из файла А не будут переписаны отсортированными во вспомогательные файлы В и С.

7. Слить данные в файл А сначала из файла В, затем из файла С. Теперь файл А содержит длинные упорядоченные серии, считаем, что данные в сериях упорядочены по возрастанию.

8. Фаза разделения включает поочередную запись серий из А в файлы В и С.

9. Фаза слияния имеет теперь следующий алгоритм:

• Считываем данные из одного и другого файлов, пока <

• После этого считываем следующую серию и так пока один из файлов не станет пустым, тогда серии другого переписываются в файл А

10. Пункты 8 и 9 повторяются пока в файл А, в результате слияния не будет переписана только одна серия.

### 2.2.1 Задание 2

Реализация алгоритма в виде блок-схем представлена на рис. 11, 12 и 13.

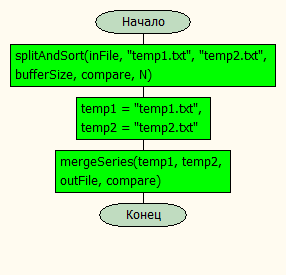


Рисунок 11 – Основная функция алгоритма

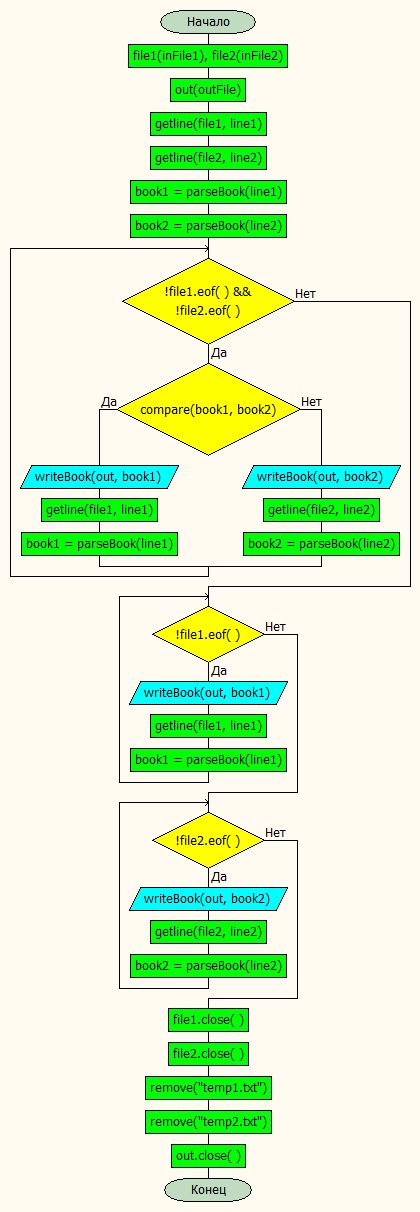


Рисунок 12 – Функция сортировки серий и записи в файл

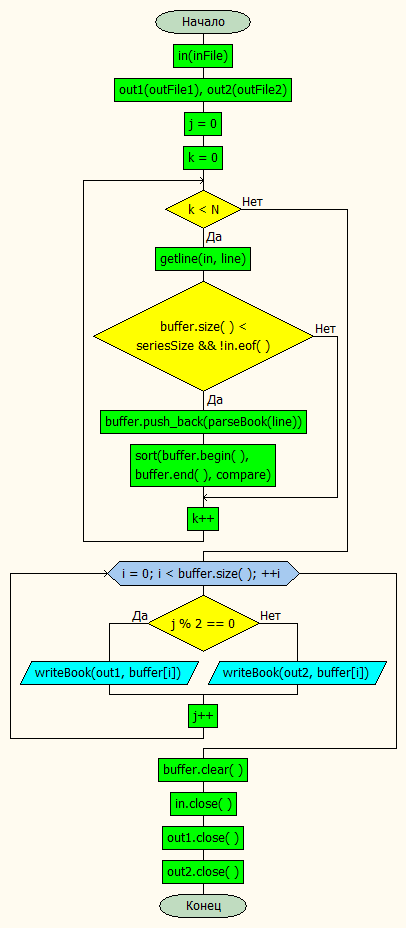


Рисунок 13 – Функция разделения данных на 2 файла

Разработаем алгоритм естественного слияния в виде функций на языке программирования С++. Код реализации представлен на рис. 14, 15 и 16.

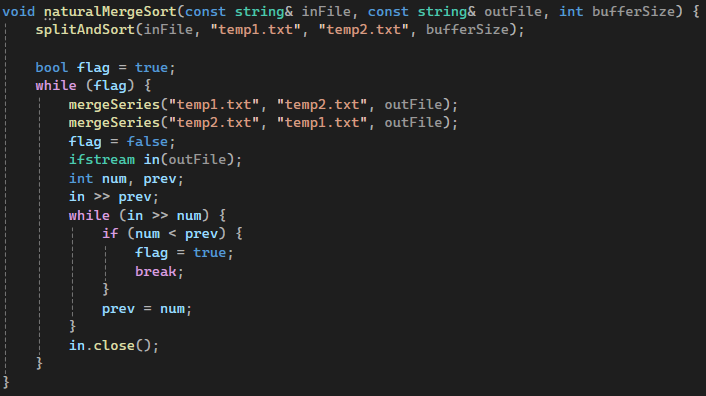


Рисунок 14 – Основная функция алгоритма естественного слияния

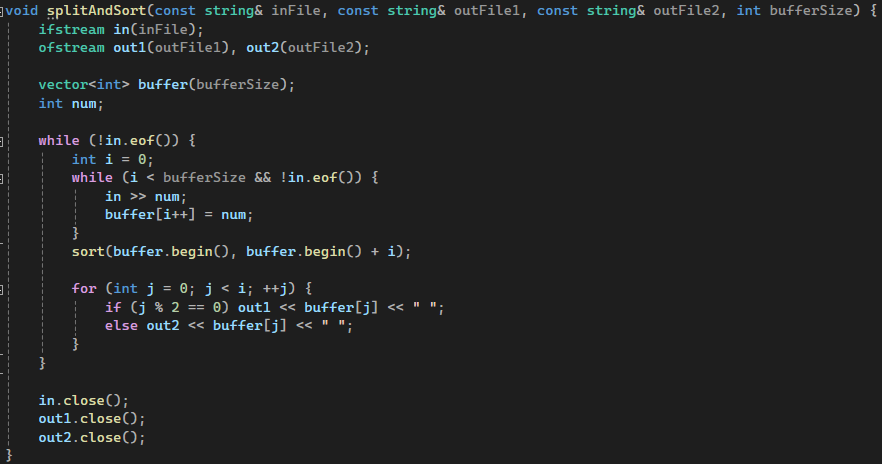


Рисунок 15 - Функция для разделения файла на серии



Рисунок 16 – Функция для слияния двух серий

Результат работы алгоритма на тестовых данных в виде файла A с записями 17 31 5 59 13 41 43 67 11 23 29 47 3 7 71 2 19 57 37 61 представлен на рис. 17.



Рисунок 17 – Результат работы алгоритма на тестовых данных

Можно сделать вывод о том, что алгоритм работает корректно.

Модернизируем алгоритм так, чтобы он работал со строками из файлов и сортировал их по ключевому элементу.

Код, реализующий алгоритм естественного слияния для файлов по ключевому элементов в виде функций представлен на рис. 18, 19.

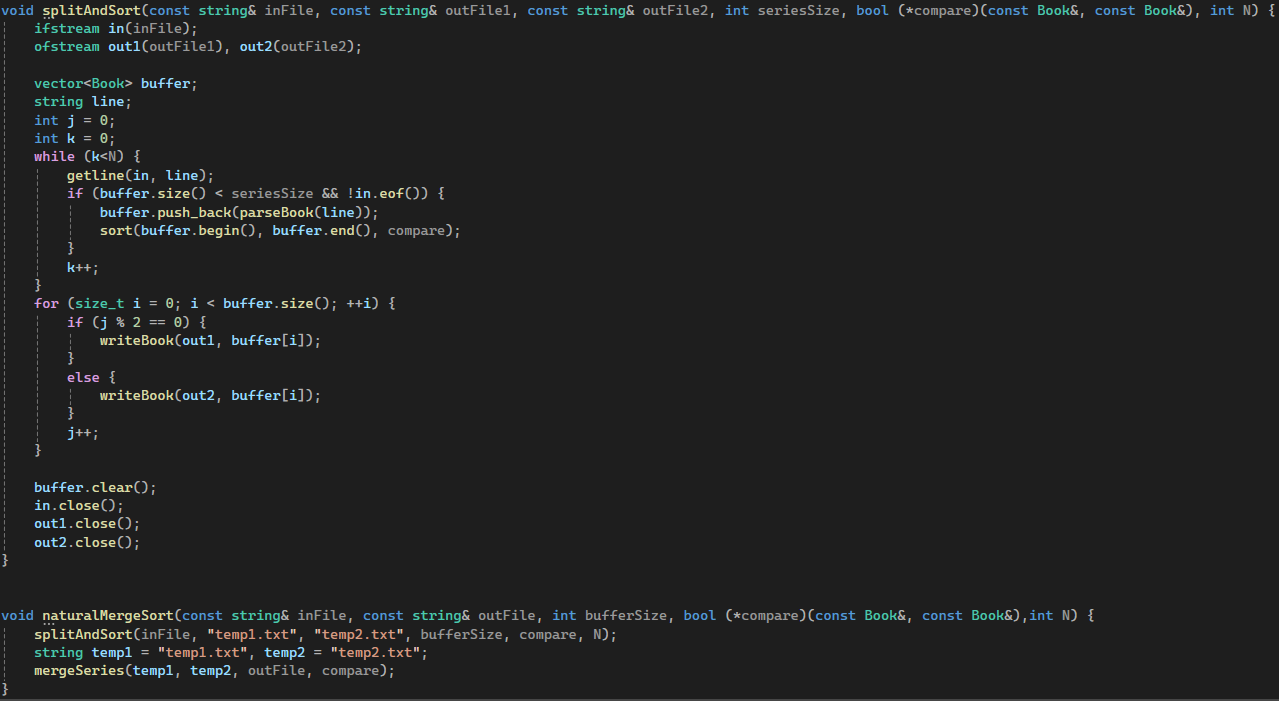


Рисунок 18 – Основная функция сортировки и разбиение файлов на временные

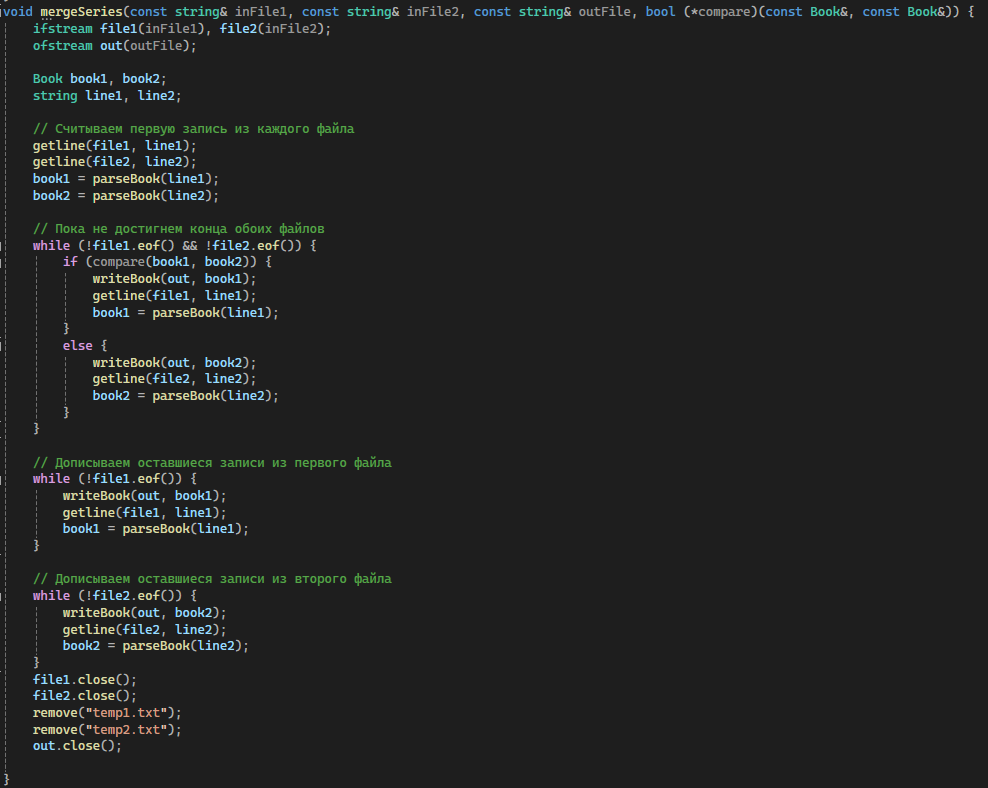


Рисунок 19 – Функция сортировки файла по сериям и записи в файл

Результаты тестирования алгоритма представлены на рис. 20.

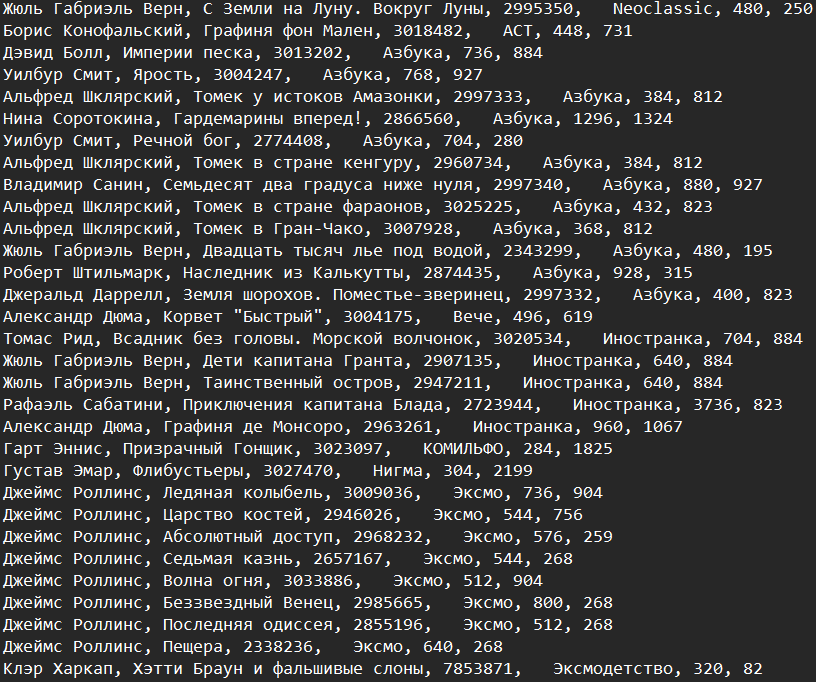


Рисунок 20 – Результаты тестирования алгоритма

Данные, дающие оценку практической сложности алгоритма, представлены в табл. 2.

Таблица 2 – Практическая сложность алгоритма естественного слияния

|  |  |
| --- | --- |
| Количество записей | Время выполнения, мс |
| 8 | 3 |
| 16 | 14 |
| 32 | 12 |

# 3 ВЫВОДЫ

В ходе выполнения работы были получены знания и умения работать с сортировками файлов, освоены приёмы сортировки данных из файлов.

По результатам тестирования алгоритмов можно убедиться в корректности их выполнения, как в случае с числовыми данными в файлах, так и со строковыми в случае наличия ключевого элемента.

По данным о практической сложности алгоритма, можно сделать два вывода:

1. Существует зависимость между временем работы алгоритмов от количества записей, с которыми они работают;
2. Алгоритм естественного слияния более эффективен, нежели алгоритм прямого слияния.

# 4 ИНФОРМАЦИОННЫЕ ИСТОЧНИКИ

1. Сартаков М.В., ПР-1.1 (Теоретическая сложность алгоритма) М., МИРЭА — Российский технологический университет – 12 с. - URL: <https://online-edu.mirea.ru/pluginfile.php?file=%2F1042738%2Fmod_assign%2Fintroattachment%2F0%2FПР1.1%20%28Теоретическая%20сложность%20алгоритма%29.pdf&amp;forcedownload=1> (дата обращения: 15.02.2024). - Режим доступа: Электронно-облачная система – Cloud MIREA РТУ МИРЭА. - Текст: электронный.
2. Рысин М.Л., Сартаков М.В., Туманова М.Б., Введение в структуры и алгоритмы обработки данных. Ч. 1 - учебное пособие, 2022, МИРЭА – Российский технологический университет. – 2022, 109с. – URL: [file:///C:/Users/borga/Downloads/Рысин%20М.Л.%20и%20др.%20Введение%20в%20структуры%20и%20алгоритмы%20обработки%20данных.%20Ч.%201%20-%20учебное%20пособие,%202022.pdf](C://Users/borga/Downloads/Рысин%20М.Л.%20и%20др.%20Введение%20в%20структуры%20и%20алгоритмы%20обработки%20данных.%20Ч.%201%20-%20учебное%20пособие,%202022.pdf) (дата обращения: 15.02.2024 ). – Режим доступа: Электронно-облачная система – Cloud MIREA РТУ МИРЭА. - Текст: электронный.