

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего

образования

**«МИРЭА - Российский технологический университет»**

РТУ МИРЭА

Институт информационных технологий

Кафедра математического обеспечения и стандартизации информационных технологий

**ОТЧЕТ**

**ПО ПРАКТИЧЕСКОЙ РАБОТЕ № 3**

***«*Хеширование и организация быстрого поиска данных*»***

по дисциплине

**«**Структуры и алгоритмы обработки данных**»**

Выполнил студент группы *ИКБО-03-22 Хохлинов Д.И.*

Принял *Сорокин А.В.*

|  |  |  |
| --- | --- | --- |
| Практическая работа выполнена | « » 2023 г. |  |
|  |  |  |
| «Зачтено» | « » 2023 г. |  |

Москва 2023

**СОДЕРЖАНИЕ**

[1 ПОСТАНОВКА ЗАДАЧИ 3](#_Toc147438816)

[2 РАЗРАБОТКА ПРИЛОЖЕНИЯ 4](#_Toc147438817)

[2.1 Структура двоичного файла 4](#_Toc147438818)

[2.2 Генерация двоичного файла 4](#_Toc147438819)

[2.3 Функционал приложения 7](#_Toc147438820)

[2.4 Код программы 11](#_Toc147438821)

[2.4 Тестирование 19](#_Toc147438822)

[2.5 Тестирование на большом объеме данных 24](#_Toc147438823)

[3 ОТВЕТЫ НА ВОПРОСЫ 31](#_Toc147438824)

[4 СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 35](#_Toc147438825)

# **1 ПОСТАНОВКА ЗАДАЧИ**

1. Разработать программу, которая использует хеш-таблицу для организации прямого доступа к записям двоичного файла, структура записи которого приведена в варианте. Разработайте класс хеш-таблицы и реализуйте методы для операций:

1) Хеширование (согласно варианту 25, используется открытая адресация с линейным пробированием);

2) Чтение записи из файла и вставка записи в таблицу (запись включает: ключ и номер записи с этим ключом в файле);

3) Удаление записи из таблицы и, соответственно, из файла;

4) Поиск записи с заданным ключом в файле, используя хеш-таблицу;

5) Выполнение рехеширования.

1. Провести тестирование программы на небольших объемах данных, введенных вручную. Разработанные тесты должны покрывать случаи возникновения коллизий и рехеширования.
2. Заполнить файл большим объемом данных. Провести исследование времени чтения записей из начала, середины и конца файла.
3. Составить отчет, отобразив в нем описание выполнения всех этапов разработки, тестирования и код всей программы со скриншотами результатов тестирования. Включить ответы на следующие вопросы:
   1. Расскажите о назначении хеш-функции.
   2. Что такое коллизия?
   3. Что такое «открытый адрес» по отношению к хеш-таблице?
   4. Как в хеш-таблице с открытым адресом реализуется коллизия?
   5. Какая проблема, может возникнуть после удаления элемента из хеш-таблицы с открытым адресом и как ее устранить?
   6. Что определяет коэффициент нагрузки в хеш-таблице?
   7. Что такое «первичный кластер» в таблице с открытым адресом?
   8. Как реализуется двойное хеширование?

# **2 РАЗРАБОТКА ПРИЛОЖЕНИЯ**

## **2.1 Структура двоичного файла**

Для выполнения задания использовалась следующая структура:

struct record {

char phone[11];

char address[100];

};

Количество памяти, занимаемое полями этой структуры:

* phone – 11 байт;
* address – 100 байт.

Итого один экземпляр занимает 111 байт.

Структура двоичного файла, используемого для хранения записей такого типа, изображена на рисунке 1.

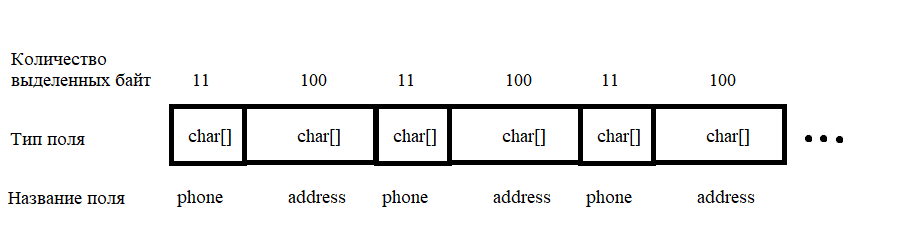


Рисунок 1 – Структура двоичного файла, используемого приложением

## **2.2 Генерация двоичного файла**

Для создания описанного двоичного файла использовались следующие программы:

* Программа для транслитерации русского текста латиницей (реализована на языке Python, листинг 1):

Листинг 1 – Код программы для транслитерации

import argparse

import sys

def create\_parser(): # command args handler

parser = argparse.ArgumentParser()

parser.add\_argument('translit')

parser.add\_argument('input')

parser.add\_argument('output')

return parser

if \_\_name\_\_ == '\_\_main\_\_':

parser = create\_parser()

args = parser.parse\_args(sys.argv[1:])

translit\_database\_path = args.translit

source\_path = args.input

output\_path = args.output

translit\_database = []

with open(translit\_database\_path, "r", encoding="utf-8") as translit\_database\_file:

lines = translit\_database\_file.readlines()

for line in lines:

split = line.split()

translit\_database.append([split[0], split[1]])

if len(translit\_database) == 0:

print("Ошибка при чтении файла с описанием замен. Процесс остановлен.")

else:

with open(source\_path, "r", encoding="utf-8") as source:

with open(output\_path, "w", encoding="utf-8") as output:

source\_lines = source.readlines()

for line in source\_lines:

for translit in translit\_database:

line = line.replace(translit[0], translit[1])

output.write(line)

print("Процесс завершен.")

* Программа для генерации двоичного файла на основе текстового (реализована на языке С++, листинг 2):

Листинг 2 – Код программы для генерации двоичного файла

#include <iostream>

#include <fstream>

#include <string>

using namespace std;

struct record {

Продолжение листинга 2

char phone[11];

char address[100];

};

int main()

{

fstream input, output;

string input\_path;

cin >> input\_path;

input = fstream(input\_path, ios::in);

output = fstream(input\_path + "\_bin.txt", ios::out | ios::binary);

record temp;

while (!input.eof())

{

string temps;

getline(input, temps);

strcpy\_s(temp.phone, temps.c\_str());

getline(input, temps);

strcpy\_s(temp.address, temps.c\_str());

output.write((char\*)&temp, sizeof(record));

}

input.close();

output.close();

}

В результате применения вышеописанных программ к текстовому файлу (листинг 3) был получен следующий двоичный файл (рисунок 2):

Листинг 3 – Исходный текстовый файл

4996008080

Moscow, Prospekt Vernadskogo, 78

3516492021

Челябинская область, город Истра, въезд Гоголя, 27

4991023923

Климентовский пер., 28, Москва, 162288

4912934834

Рязанская область, город Серебряные Пруды, въезд Ленина, 01

4920493923

Владимирская область, город Балашиха, ул. Славы, 48

3837461212

Новосибирская область, город Домодедово, шоссе Ленина, 90

3439992283

Свердловская область, город Подольск, наб. Бухарестская, 05

4751923845

Тамбовская область, город Павловский Посад, спуск 1905 года, 93

4739328456

Воронежская область, город Дмитров, въезд Домодедовская, 76

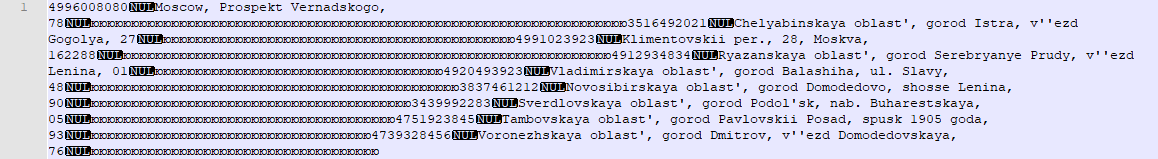


Рисунок 2 – Полученный двоичный файл

## **2.3 Функционал приложения**

Для выполнения задания использовались следующие функции и модули:

1. Структура record:
   1. Конструктор по умолчанию: заполняет структуру данными по умолчанию;
   2. Конструктор с параметрами: заполняет структуру данными из аргументов;
   3. Оператор ==: если все поля структур равны друг другу, то возвращает «истину», иначе – «ложь»;
   4. Оператор !=: возвращает инверсию результата сравнения структур.
2. Функция getHash. Принимает в качестве аргумента указатель на переменную типа record. Вычисляет сумму цифр, записанных в поле phone переданной структуры, в степени, равной индексу цифры в поле, затем возвращает остаток от деления этой суммы на размер хэш-таблицы.
3. Структура hashtable\_entry:
   1. Поля:
      1. open – тип: bool;
      2. wasCollision – тип: bool;
      3. hash – тип: int;
      4. record\_number – тип: int;
      5. key – тип: массив из 11 элементов char;
   2. Конструктор с параметром hash: присваивает соответствующему полю значение hash;
   3. Деструктор: пустой.
4. Класс hashtable:
   1. Поля:
      1. entries – динамический массив типа указателей на hashtable\_entry;
      2. used\_entries – тип: int;
   2. Конструктор по умолчанию: пустой;
   3. Конструктор с параметром size: создает size записей в поле entries;
   4. Деструктор: очищает память, выделенную под массив entries;
   5. Метод getUsage: возвращает результат деления значения поля used\_entries на размер таблицы table\_size;
   6. Метод show: выводит на экран текущее состояние таблицы;
   7. Метод extend\_table: пока значение, возвращаемое методом getUsage таблицы, выше 0,75, увеличивает значение table\_size в два раза. Затем добавляет новые записи в массив entries и вызывает метод rehash с параметрами false, -1;
   8. Метод rehash. Принимает в качестве параметров режим работы mode и целое число pos. Открывает файл, с которым проводится работа; если его не удалось открыть, выводит соответствующее сообщение и завершает работу приложения. В противном случае сохраняет имеющиеся данные в таблице по следующему алгоритму:
      1. Объявить динамический целочисленный массив records;
      2. Сохранить номер записи в очередной строке таблицы как rn;
      3. Если rn не равно -1, то:
         1. Если mode – «ложь», то добавить rn в records. Если mode – «истина», то:
            1. Если rn равно pos, то ничего не происходит;
            2. Если rn больше pos, то добавить rn-1 в records;
            3. Если rn меньше pos, то добавить rn в records.
         2. Очистить данные в строке таблицы.
      4. Если конец таблицы не достигнут, то переход к п.2.

После сохранения данных производится сортировка records с помощью функции std::sort. Затем из файла считываются записи, указанные в массиве records, и записываются в таблицу с помощью addData.

* 1. Метод addData. Принимает в качестве аргументов указатель data на структуру record и целое число record\_number. Если переданная структура уже есть в таблице, выводит соответствующее сообщение. В противном случае вычисляет хэш hash переданной структуры. Если заполнено более 75% таблицы, вызывает метод extend\_table. Объявляет переменные c = TABLE\_STEP = 3 (шаг пробирования) и i = 0. Затем пытается записать данные о структуре в строку таблицы с хэшем hash + c \* i; если эта ячейка уже занята, помечает ее как ячейку, в которой наблюдалась коллизия и увеличивает i на 1, после чего повторяет попытку. В противном случае записывает данные в открытую ячейку, помечает ее как закрытую и увеличивает поле used\_entries на 1;
  2. Метод removeData. Принимает в качестве аргумента указатель data на структуру record. Если переданная структура не найдена, выводит соответствующее сообщение. В противном случае перезаписывает файл так, чтобы остались все записи, кроме найденной, и вызывает метод rehash с параметрами true, pos, где pos – значение, возвращенное методом findData с аргументом data->phone;
  3. Метод findData. Принимает в качестве аргумента ключ phone. Вычисляет хэш hash структуры с ключом phone, объявляет счетчик итераций iterations = 0. Затем действует по следующему алгоритму:
     1. Если ячейка с хэшем hash открыта, то:
        1. Если в этой ячейке не возникало коллизий, то возвращает -1;
        2. В противном случае увеличивает hash на величину шага (TABLE\_STEP = 3);
     2. В противном случае:
        1. Если ключ текущей ячейки совпадает с исходным, то возвращает номер записи, записанной в этой ячейке;
        2. В противном случае увеличивает hash на величину шага (TABLE\_STEP = 3);
     3. Если hash больше или равен размеру таблицы, то уменьшает hash на table\_size;
     4. Если количество итераций превысило двойной размер таблицы - аварийно завершить работу, так как это свидетельствует о возникших ошибках при работе с таблицей. Если нет, то переходит к п.1.
  4. Меню модуля process. Подготавливает таблицу к работе (изначальный размер таблицы = table\_size = 5). Затем в цикле предлагает пользователю одно из доступных действий. После выбора пользователем действия «выход» завершает работу.

## **2.4 Код программы**

Запишем код программы на языке С++ (листинги 4-6):

Листинг 4 – Hashtable.h

#pragma once

#ifndef \_\_hashtable\_h\_\_

#define \_\_hashtable\_h\_\_

#include <iostream>

#include <algorithm>

#include <fstream>

#include <string>

#include <vector>

#include <conio.h>

using namespace std;

namespace Hashtable {

struct record {

char phone[11];

char address[100];

record();

record(char\* phone, char\* address);

Продолжение листинга 4

bool operator == (record\* right);

bool operator != (record\* right);

};

int getHash(record\* data);

struct hashtable\_entry {

bool open = true;

bool wasCollision = false;

int hash = 0;

int record\_number = -1;

char key[11];

hashtable\_entry(int hash);

~hashtable\_entry();

};

class hashtable {

vector<hashtable\_entry\*> entries;

int used\_entries = 0;

public:

float getUsage();

void show();

bool addData(record\* data, int record\_number);

bool removeData(record\* data);

int findData(char\* phone);

void process();

void rehash(bool mode, int pos);

void extend\_table();

hashtable();

hashtable(int size);

~hashtable();

};

}

#endif

Листинг 5 – Hashtable.cpp

#include "Hashtable.h"

int table\_size = 5;

const int TABLE\_STEP = 3;

string path;

bool errors = false;

Hashtable::record::record()

{

for (int i = 0; i < 11; i++)

{

this->phone[i] = 0;

this->address[i] = 0;

}

for (int i = 11; i < 100; i++)

{

this->address[i] = 0;

Продолжение листинга 5

}

strcpy\_s(this->phone, "1234567890");

strcpy\_s(this->address, "Generic address");

}

Hashtable::record::record(char\* phone, char\* address)

{

for (int i = 0; i < 11; i++)

{

this->phone[i] = 0;

this->address[i] = 0;

}

for (int i = 11; i < 100; i++)

{

this->address[i] = 0;

}

strcpy\_s(this->phone, phone);

strcpy\_s(this->address, address);

}

bool Hashtable::record::operator ==(record\* right)

{

if (strcmp(this->phone, right->phone)) return false;

if (strcmp(this->address, right->address)) return false;

return true;

}

bool Hashtable::record::operator!=(record\* right)

{

return !(\*this == right);

}

int Hashtable::getHash(record\* data)

{

unsigned int temp = 0;

for (int i = 0; i < 11; i++)

{

temp += pow((abs((int)(data->phone[i])-48)), (i+1));

}

return temp % table\_size;

}

Hashtable::hashtable\_entry::hashtable\_entry(int hash) : hash(hash)

{

for (int i = 0; i < 11; i++)

{

this->key[i] = 0;

}

}

Hashtable::hashtable\_entry::~hashtable\_entry()

{

}

Продолжение листинга 5

Hashtable::hashtable::hashtable()

{

}

Hashtable::hashtable::hashtable(int size)

{

for (int i = 0; i < size; i++)

{

hashtable\_entry\* temp = new hashtable\_entry(i);

this->entries.push\_back(temp);

}

}

Hashtable::hashtable::~hashtable()

{

while (this->entries.size() > 0)

this->entries.pop\_back();

}

float Hashtable::hashtable::getUsage()

{

return float(this->used\_entries) / table\_size;

}

void Hashtable::hashtable::extend\_table() {

int current\_size = table\_size;

while (this->getUsage() >= 0.75) table\_size \*= 2;

for (int i = current\_size; i < table\_size; i++)

{

hashtable\_entry\* temp = new hashtable\_entry(i);

this->entries.push\_back(temp);

}

this->rehash(false, -1);

}

void Hashtable::hashtable::rehash(bool mode, int pos)

{

this->used\_entries = 0;

fstream datafile = fstream(path, ios::in | ios::binary);

if (datafile)

{

record temp;

int current\_record = 0;

vector<int> records;

for (int i = 0; i < table\_size; i++)

{

int rn = this->entries.at(i)->record\_number;

if (rn != -1)

{

if (mode)

{

if (rn == pos) {}

else if (rn > pos)

records.push\_back(rn-1);

Продолжение листинга 5

else

records.push\_back(rn);

}

else

records.push\_back(rn);

this->entries.at(i)->record\_number = -1;

this->entries.at(i)->open = true;

this->entries.at(i)->wasCollision = false;

strcpy\_s(this->entries.at(i)->key, "");

}

}

sort(records.begin(), records.end());

while (!datafile.eof())

{

if (current\_record == records.size()) break;

int start = 0;

if (current\_record) start = records[current\_record-1];

int end = records[current\_record];

for (int i = start; i < end; i++)

datafile.read((char\*)&temp, sizeof(record));

this->addData(&temp, records[current\_record]);

current\_record++;

}

}

else {

cout << "Ошибка при чтении файла." << endl;

errors = true;

}

}

bool Hashtable::hashtable::addData(record\* data, int record\_number)

{

if (this->findData(data->phone) > 0)

{

cout << "Ошибка: эта запись уже есть в таблице." << endl;

return false;

}

int hash = getHash(data);

int c = TABLE\_STEP;

int i = 0;

if (this->getUsage() >= 0.75) this->extend\_table();

while (this->entries.at(hash + c \* i)->open == false)

{

this->entries.at(hash + c \* i)->wasCollision = true;

i++;

if (hash + c \* i >= table\_size) hash -= table\_size;

}

this->entries.at(hash + c \* i)->record\_number = record\_number;

this->entries.at(hash + c \* i)->open = false;

strcpy\_s(this->entries.at(hash + c \* i)->key, data->phone);

this->used\_entries++;

return true;

}

Продолжение листинга 5

bool Hashtable::hashtable::removeData(record\* data)

{

int pos = this->findData(data->phone);

if (pos < 0)

{

cout << "Элемент с ключом " << data->phone << " не найден в хэш-таблице" << endl;

return false;

}

else

{

fstream tempfile("temp.txt", ios::out | ios::binary);

fstream source(path, ios::in | ios::binary);

record temp;

int i = 1;

while (source.read((char\*)&temp, sizeof(record)))

{

if (i != pos)

{

tempfile.write((char\*)&temp, sizeof(record));

}

i++;

}

source.close();

tempfile.close();

tempfile = fstream("temp.txt", ios::in | ios::binary);

source = fstream(path, ios::out | ios::binary);

while (tempfile.read((char\*)&temp, sizeof(record)))

{

source.write((char\*)&temp, sizeof(record));

}

source.close();

tempfile.close();

this->rehash(true, pos);

}

return true;

}

int Hashtable::hashtable::findData(char\* phone)

{

record data = record(phone, phone);

int hash = getHash(&data);

int iterations = 0;

while (iterations < 2 \* table\_size) //для предотвращения зацикливания

{

if (this->entries.at(hash)->open)

{

if (this->entries.at(hash)->wasCollision == false)

{

return -1;

}

}

else

{

Продолжение листинга 5

if (!strcmp(this->entries.at(hash)->key, data.phone))

{

return this->entries.at(hash)->record\_number;

}

}

hash += TABLE\_STEP;

if (hash >= table\_size) hash -= table\_size;

iterations++;

}

cout << "Ошибка при поиске в хэш-таблице." << endl;

errors = true;

return -2;

}

void Hashtable::hashtable::show()

{

for (int i = 0; i < this->entries.size(); i++)

{

hashtable\_entry\* temp = this->entries.at(i);

cout << "hash: " << temp->hash << ", open: " << temp->open << ", was collision: " << temp->wasCollision << ", record number: " << temp->record\_number << ", key: " << temp->key << endl;

}

}

void Hashtable::hashtable::process()

{

path = "input.txt";

fstream input;

for (int i = 0; i < table\_size; i++)

{

hashtable\_entry\* temp = new hashtable\_entry(i);

this->entries.push\_back(temp);

}

record temp;

int i = 1;

bool running = true;

char choose = ' ';

int record\_number, pos;

char key[11];

strcpy\_s(key, "");

while (running)

{

cout << "Выберите действие:" << endl

<< "[Q] - выход" << endl

<< "[A] - добавить запись в таблицу" << endl

<< "[R] - удалить запись из таблицы и из файла" << endl

<< "[F] - найти запись в таблице" << endl

<< "[S] - показать хэш-таблицу" << endl;

choose = \_getch();

switch (tolower(choose))

{

case 'q':

running = false;

Продолжение листинга 5

break;

case 's':

cout << "Хэш-таблица (использовано полей: " << this->used\_entries << "/" << table\_size << "):" << endl;

this->show();

break;

case 'a':

cout << "Введите номер записи из файла, который необходимо вставить в таблицу: ";

if (cin >> record\_number)

{

input = fstream(path, ios::in | ios::binary);

i = 1;

while (input.read((char\*)&temp, sizeof(record)))

{

if (i == record\_number)

break;

i++;

}

if (i != record\_number)

{

cout << "Ошибка при чтении записи №" << record\_number << endl;

}

else

{

this->addData(&temp, record\_number);

}

input.close();

}

else

{

cout << "Ошибка при вводе значения." << endl;

}

break;

case 'r':

cout << "Введите номер телефона, запись с которым нужно удалить: ";

cin >> key;

if (this->removeData(&record(key, key)))

{

cout << "Запись с ключом " << key << " удалена успешно." << endl;

}

else

{

cout << "Ошибка при удалении записи с ключом " << key << endl;

}

break;

case 'f':

cout << "Введите номер телефона, запись с которым нужно найти: ";

cin >> key;

pos = this->findData(key);

Продолжение листинга 5

if (pos > 0)

{

cout << "Запись с ключом " << key << " найдена на позиции " << pos << endl;

}

else

{

cout << "Запись с ключом " << key << " не найдена" << endl;

}

break;

default:

cout << "Неизвестное действие." << endl;

}

if (errors)

{

cout << "Были обнаружены ошибки, которые могут влиять на работу программы. Процесс прекращен." << endl;

running = false;

}

if (running)

{

system("pause");

system("cls");

}

}

}

Листинг 6 – main.cpp

#include <iostream>

#include <locale.h>

#include <fstream>

#include "Hashtable.h"

using namespace std;

int main()

{

setlocale(LC\_ALL, "ru");

Hashtable::hashtable table = Hashtable::hashtable();

table.process();

}

## **2.4 Тестирование**

Проведем тестирование созданного приложения на небольшом двоичном файле (рисунок 2). Ход и результаты тестирования приведены на рисунках 3-13:

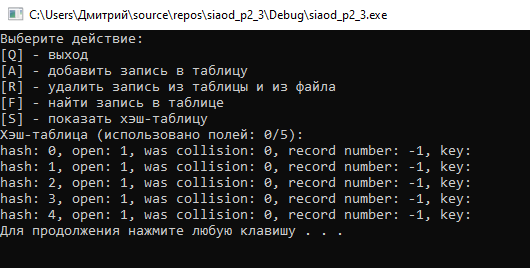


Рисунок 3 – Хэш-таблица в стартовом состоянии

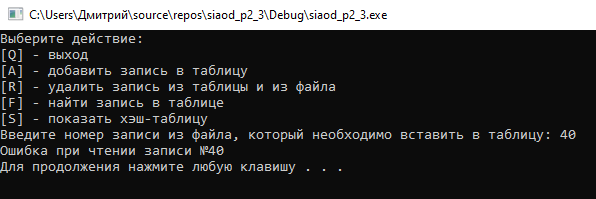


Рисунок 4 – Попытка прочитать запись с несуществующим номером

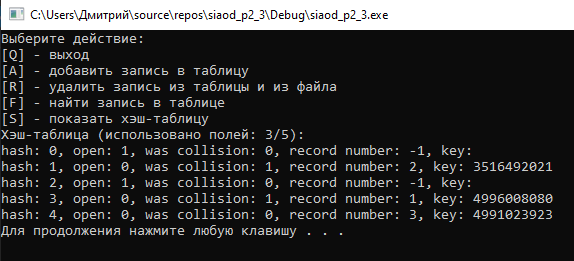


Рисунок 5 – Хэш-таблица после чтения первых 3 записей

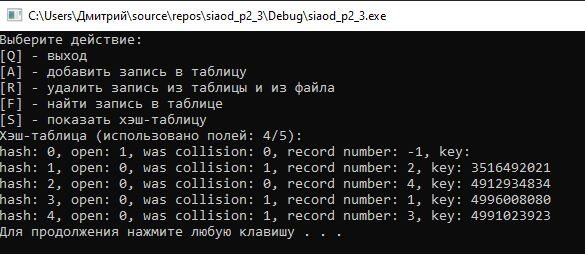


Рисунок 6 – Хэш-таблица после чтения четвертой записи, при котором произошла коллизия

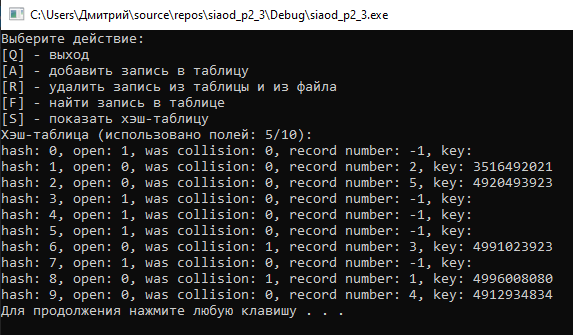


Рисунок 7 – Хэш-таблица после чтения пятой записи. Произошло рехэширование из-за высокой загруженности таблицы (0.8)

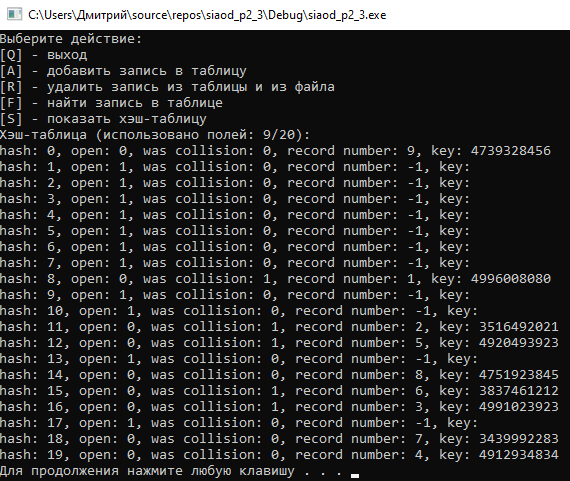


Рисунок 8 – Хэш-таблица после чтения 6-9 записей. При чтении 9 записи выполнялось рехэширование

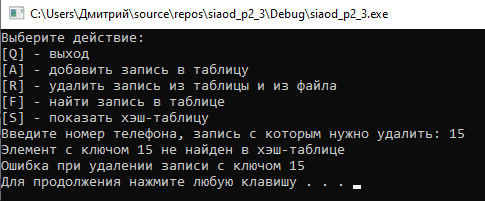


Рисунок 9 – Попытка удалить из таблицы несуществующую запись

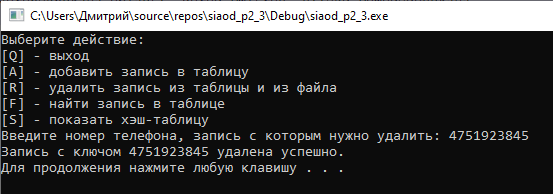


Рисунок 10 – Удаление из таблицы существующей записи

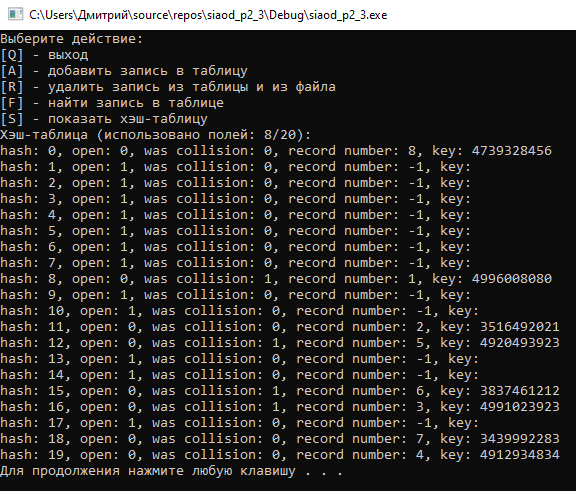


Рисунок 11 – Хэш-таблица после удаления, показанного на рисунке 10. Производилось рехэширование, связанное с изменением номеров записей

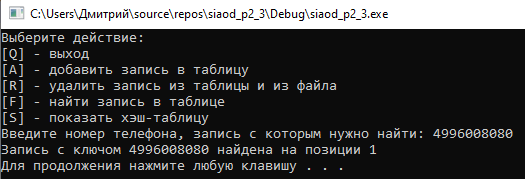


Рисунок 12 – Поиск существующей записи

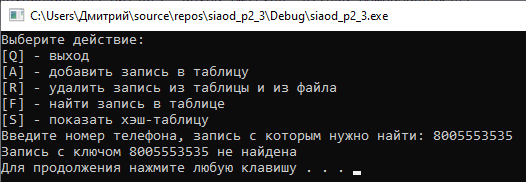


Рисунок 13 – Попытка поиска несуществующей записи

## **2.5 Тестирование на большом объеме данных**

При тестировании работы программы с большим количеством данных использовался двоичный файл, полученный в результате манипуляций, описанных в пункте 2.2, над текстовым файлом большого объема (1000 записей).

Проведем модификацию используемой программы:

* Изменим метод process так, чтобы перед началом работы с таблицей она уже была заполнена всеми записями из файла;
* В методе process добавим новое доступное пользователю действие: прочитать запись, найденную по ключу, из файла и вывести ее на экран. При этом замеряется время поиска записи в хэш-таблице;
* Для уменьшения числа рехэширований установим начальный размер таблицы равным 1000.

Листинг 7 – Измененный метод process в файле Hashtable.cpp

void Hashtable::hashtable::process()

{

path = "big input.txt";

fstream input;

for (int i = 0; i < table\_size; i++)

{

hashtable\_entry\* temp = new hashtable\_entry(i);

this->entries.push\_back(temp);

}

record temp;

int i = 1;

input = fstream(path, ios::in | ios::binary);

while (input.read((char\*)&temp, sizeof(record)))

{

this->addData(&temp, i);

i++;

}

input.close();

i = 1;

auto start = chrono::steady\_clock::now();

auto end = chrono::steady\_clock::now();

bool running = true;

char choose = ' ';

int record\_number, pos;

char key[11];

strcpy\_s(key, "");

while (running)

{

cout << "Выберите действие:" << endl

<< "[Q] - выход" << endl

<< "[A] - добавить запись в таблицу" << endl

<< "[R] - удалить запись из таблицы и из файла" << endl

<< "[F] - найти запись в таблице" << endl

<< "[O] - найти запись в таблице и вывести ее на экран" << endl

<< "[S] - показать хэш-таблицу" << endl;

choose = \_getch();

switch (tolower(choose))

{

case 'q':

running = false;

break;

case 's':

cout << "Хэш-таблица (использовано полей: " << this->used\_entries << "/" << table\_size << "):" << endl;

this->show();

break;

case 'a':

cout << "Введите номер записи из файла, который необходимо вставить в таблицу: ";

if (cin >> record\_number)

{

input = fstream(path, ios::in | ios::binary);

i = 1;

Продолжение листинга 7

while (input.read((char\*)&temp, sizeof(record)))

{

if (i == record\_number)

break;

i++;

}

if (i != record\_number)

{

cout << "Ошибка при чтении записи №" << record\_number << endl;

}

else

{

this->addData(&temp, record\_number);

}

input.close();

}

else

{

cout << "Ошибка при вводе значения." << endl;

}

break;

case 'r':

cout << "Введите номер телефона, запись с которым нужно удалить: ";

cin >> key;

if (this->removeData(&record(key, key)))

{

cout << "Запись с ключом " << key << " удалена успешно." << endl;

}

else

{

cout << "Ошибка при удалении записи с ключом " << key << endl;

}

break;

case 'f':

cout << "Введите номер телефона, запись с которым нужно найти: ";

cin >> key;

pos = this->findData(key);

if (pos > 0)

{

cout << "Запись с ключом " << key << " найдена на позиции " << pos << endl;

}

else

{

cout << "Запись с ключом " << key << " не найдена" << endl;

}

break;

case 'o':

Продолжение листинга 7

cout << "Введите номер телефона, запись с которым нужно найти: ";

cin >> key;

start = chrono::steady\_clock::now();

pos = this->findData(key);

end = chrono::steady\_clock::now();

if (pos > 0)

{

cout << "Запись с ключом " << key << " найдена на позиции " << pos << endl;

input = fstream(path, ios::in | ios::binary);

i = 1;

while (input.read((char\*)&temp, sizeof(record)))

{

if (i == pos)

break;

i++;

}

}

else

{

cout << "Запись с ключом " << key << " не найдена" << endl;

}

if (pos > 0)

{

cout << "Найденная запись: " << temp.phone << "; " << temp.address << endl;

}

cout << "Время поиска в хэш-таблице: " << chrono::duration\_cast<chrono::microseconds> (end - start).count() << " мкс" << endl;

break;

default:

cout << "Неизвестное действие." << endl;

}

if (errors)

{

cout << "Были обнаружены ошибки, которые могут влиять на работу программы. Процесс прекращен." << endl;

running = false;

}

if (running)

{

system("pause");

system("cls");

}

}

}

Проведем тестирование модифицированной программы. В ходе этого тестирования будут произведены поиски записи в начале, середине и конце файла, а также поиск несуществующей записи. Для уменьшения погрешности измерения тесты будут выполнены по 5 раз, а в качестве результата будет взято среднее значение. Результаты приведены в таблице 1, примеры вывода представлены на рисунках 14-17:

Таблица 1 – Время поиска различных записей с помощью хэш-таблицы

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Искомый объект | Время поиска объекта, мкс | | | | | |
| № измерения | | | | | Среднее значение |
| 1 | 2 | 3 | 4 | 5 |
| Запись в начале файла (ключ: 4016588676) | 6 | 7 | 7 | 6 | 7 | 6.6 |
| Запись в середине файла (ключ: 2387214004) | 9 | 7 | 9 | 10 | 10 | 9 |
| Запись в конце файла (ключ: 4758479848) | 15 | 15 | 15 | 15 | 16 | 15.2 |
| Запись, не имеющаяся в файле (ключ: 8005553535) | 15 | 15 | 16 | 15 | 15 | 15.2 |

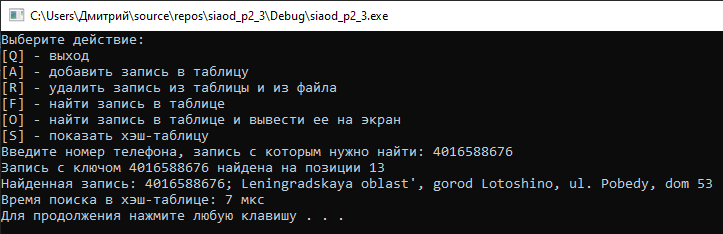


Рисунок 14 – Поиск записи в начале файла

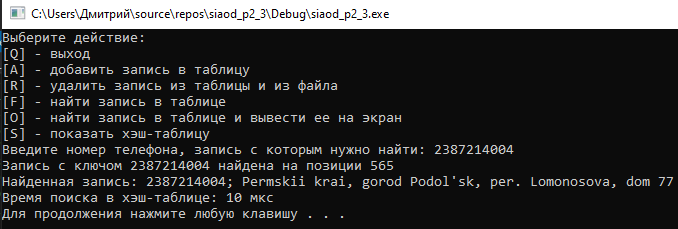


Рисунок 15 – Поиск записи в середине файла

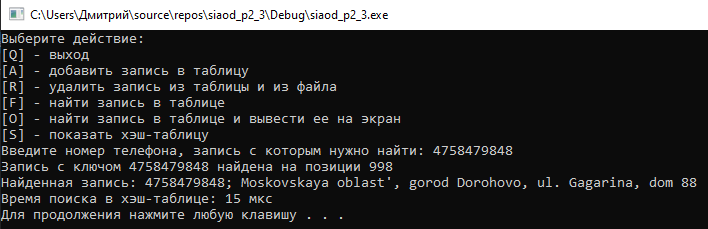


Рисунок 16 – Поиск записи в конце файла

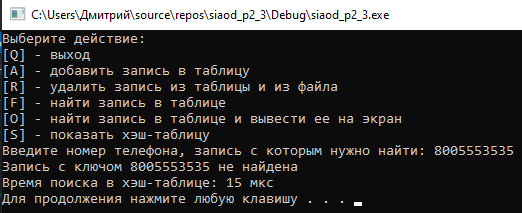


Рисунок 17 – Поиск несуществующей записи

По результатам тестирования видно, что время выполнения поиска в хэш-таблице отлично от О(1), что связано с большим количеством коллизий в хэш-таблице.

# **3 ОТВЕТЫ НА ВОПРОСЫ**

Вопрос 1 - Расскажите о назначении хеш-функции.

Хэш-функция – это математический алгоритм, преобразовывающий произвольные данные в число или строку фиксированной длины. При условии использования одного и того же алгоритма длина строки постоянна и не зависит от входных данных. Хэш-функции используются для шифрования паролей, создания хэш-таблиц, служащих для ускорения доступа к данным, и т.д.

Значение хэш-функции для одного и того же набора входных данных всегда одно и то же, для разных наборов входных данных значения хэш-функции могут совпадать.

Вопрос 2 - Что такое коллизия?

Коллизия – ситуация, при которой разным входным данным соответствует одно и то же значение хэш-функции. В общем случае коллизии неизбежны, но в оптимальных хэш-функциях их количество сведено к минимуму.

Вопрос 3 - Что такое «открытый адрес» по отношению к хеш-таблице?

«Открытый адрес» - строка хэш-таблицы, которая не занята никаким значением. Эта строка может быть перезаписана в процессе работы, тогда адрес станет «закрытым».

Вопрос 4 - Как в хеш-таблице с открытым адресом реализуется коллизия?

При возникновении коллизии в таблицах с открытым адресом происходит вычисление другого адреса внутри таблицы по некоторой функции, зависящей от текущего хэша и счетчика i=0, увеличивающемся на 1 при каждой коллизии. Таким образом, при попадании в «закрытую» ячейку происходит переход в другую, «связанную» с ней ячейку.

Вопрос 5 - Какая проблема, может возникнуть после удаления элемента из хеш-таблицы с открытым адресом и как ее устранить?

После удаления элемента из хэш-таблицы с открытым адресом может возникнуть ситуация, представленная на рисунке 18.

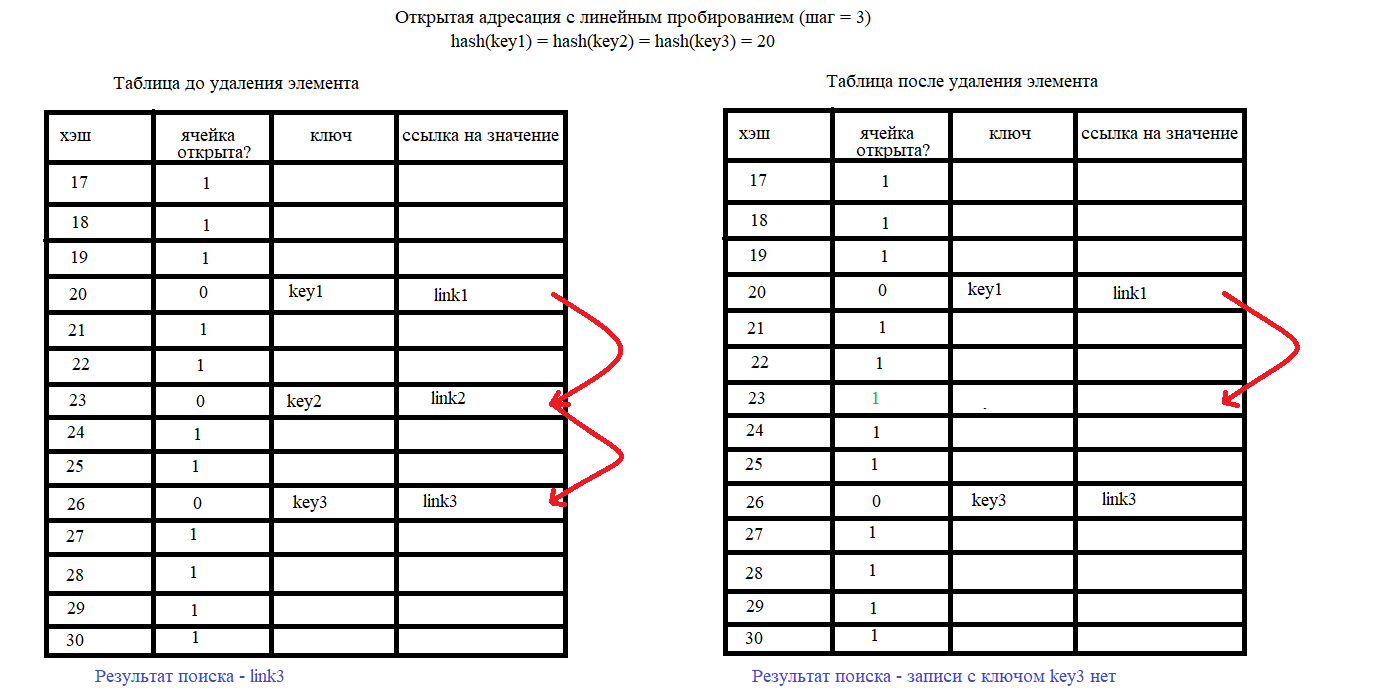


Рисунок 18 – Возможная проблема при удалении элемента из хэш-таблицы

До удаления записи с ключом key2 поиск элемента с ключом key3 проводился по следующей схеме:

1. Вычислить хэш элемента (20);
2. Ячейка с хэшем 20 закрыта;
3. Ключ key1 не равен ключу key3 – искомый элемент не найден, выполнить смещение на 3 ячейки;
4. Ячейка с хэшем 23 закрыта;
5. Ключ key2 не равен ключу key3 – искомый элемент не найден, выполнить смещение на 3 ячейки;
6. Ячейка с хэшем 26 закрыта;
7. Ключ key3 равен ключу key3 – искомый элемент найден.

После удаления записи с ключом key2 поиск элемента с ключом key3 будет проходить так:

1. Вычислить хэш элемента (20);
2. Ячейка с хэшем 20 закрыта;
3. Ключ key1 не равен ключу key3 – искомый элемент не найден, выполнить смещение на 3 ячейки;
4. Ячейка с хэшем 23 открыта – искомый элемент не содержится в таблице.

То есть, получаем, что после удаления элемента key2 мы не сможем найти элемент key3 по таблице. Для решения этой проблемы нужно ввести дополнительное поле, показывающее, возникали ли коллизии при записи в эту ячейку (рисунок 19):

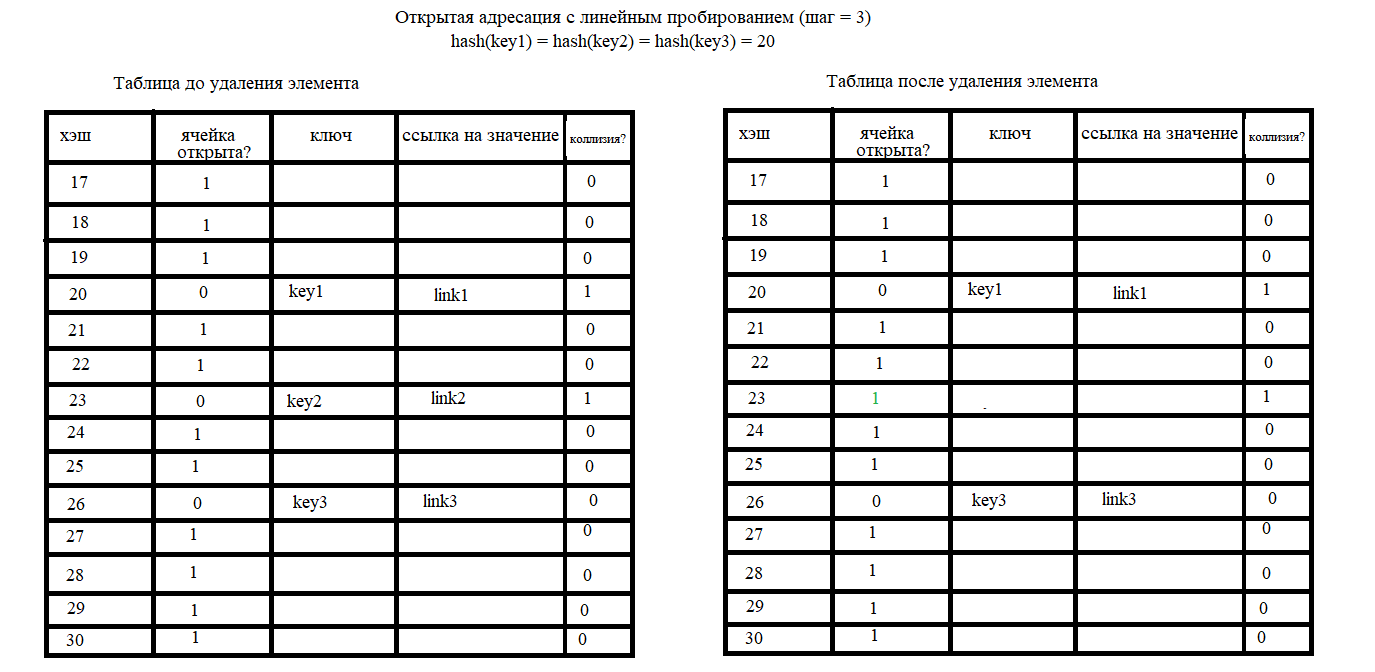


Рисунок 19 – Хэш-таблица с дополнительным полем

Также нужно изменить алгоритм поиска: если очередная ячейка открыта, то нужно прочитать значение поля, отвечающего за коллизии: если в этой ячейке случались коллизии, то продолжаем поиск, так как искомый элемент может быть расположен за удаленным элементом; если нет, то завершаем поиск.

Вопрос 6 - Что определяет коэффициент нагрузки в хеш-таблице?

Коэффициент нагрузки хэш-таблицы рассчитывается как , где u – число записей в таблице, s – размер таблицы, и определяет степень загруженности хэш-таблицы. Если w > 0.75, то следует увеличить размер таблицы в 2 раза и провести рехеширование.

Вопрос 7 - Что такое «первичный кластер» в таблице с открытым адресом?

«Первичный кластер» - ситуация, при которой записи занимают соседние строки хэш-таблицы, образуя большое количество коллизий. Это может быть вызвано неудачным выбором хэш-функции или шага пробирования. Чтобы избежать данной ситуации, нужно подобрать такой шаг пробирования c, чтобы числа c и N – размер таблицы – были взаимно простыми, а с – не очень малым числом.

Вопрос 8 - Как реализуется двойное хеширование?

При реализации двойного хеширования используются две хэш-функции f(x) и h(x), которые имеют схожее, но не одинаковое действие. При этом функция f(x) определяет хэш аргумента, h(x) – смещение при возникновении коллизии. Эта схема позволяет уменьшить вероятность возникновения «первичного кластера» при заполнении таблицы.

# **4 СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Задание на самостоятельную работу: <https://online-edu.mirea.ru/pluginfile.php?file=%2F1144079%2Fmod_assign%2Fintroattachment%2F0%2F%D0%A1%D0%B8%D0%90%D0%9E%D0%94%20%D0%A1%D0%B0%D0%BC%D0%BE%D1%81%D1%82%D0%BE%D1%8F%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F%20%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%B0%203%20%28%D1%85%D0%B5%D1%88%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5%29.pdf&amp;forcedownload=1>, дата обращения: 05.10.23
2. Структуры и алгоритмы обработки данных – Лекция 2.1: <https://online-edu.mirea.ru/pluginfile.php?file=%2F1126225%2Fmod_folder%2Fcontent%2F0%2F%D0%A1%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D1%8B%20%D0%B8%20%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC%D1%8B%20%D0%BE%D0%B1%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%BA%D0%B8%20%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85_%D0%9B%D0%B5%D0%BA%D1%86%D0%B8%D1%8F_2.1.pptx&forcedownload=1>, дата обращения: 05.10.23
3. Методические указания к практической работе: <https://online-edu.mirea.ru/pluginfile.php?file=%2F1126230%2Fmod_folder%2Fcontent%2F0%2F%D0%A1%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D1%8B%20%D0%B8%20%D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC%D1%8B%20%D0%BE%D0%B1%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%BA%D0%B8%20%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85_%D0%9F%D1%80%D0%B0%D0%BA%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F%20%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%B0_2.3%20%28%D0%A5%D0%B5%D1%88%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5%29.docx&forcedownload=1>, дата обращения: 05.10.23