

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА - Российский технологический университет"

РТУ МИРЭА

Институт информационных технологий (ИТ)
Кафедра математического обеспечения и стандартизации информационных технологий
(МОСИТ)

ОТЧЁТ ПО ПРАКТИЧЕСКОМУ ЗАДАНИЮ №4 по дисциплине «Структуры и алгоритмы обработки данных»

Тема: «Сбалансированные деревья поиска (СДП) и их применение для поиска данных

в файле»

Отчет представлен к рассмотрению:
Студентка группы ИНБО-01-20 «06» ноября 2021 г. (подпись)

Преподаватель «06» ноября 2021 г. (подпись)

Сорокин А.В.

СОДЕРЖАНИЕ

Цель работы	3
Задание 1	3
Задание 2	10
Задание 3	20
Вывод	32
Список информационных источников	33

Цель работы

Получение навыков в разработке и реализации алгоритмов управления бинарным деревом поиска и сбалансированными бинарными деревьями поиска (красно-чёрными деревьями), в применении файловых потоков прямого доступа к данным файла, в применении сбалансированного дерева поиска для прямого доступа к записям файла.

Задание 1

Постановка задачи.

Разработать приложение, которое использует бинарное дерево поиска (БДП) для поиска записи с ключом в файле.

Вариант 11.

Структура элемента множества (ключ – подчеркнутое поле) Железнодорожная справка: номер поезда, пункт отправления, пункт назначения, время отправления

Дано:

Класс «Бинарное дерево поиска». Методы: включение элемента в дерево, поиск ключа в дереве, удаление ключа из дерева, отображение дерева.

Класс управления файлом. Методы: создание двоичного файла записей фиксированной длины из заранее подготовленных данных в текстовом файле; поиск записи в файле с использованием БДП.

Результат.

Приложение, выполняющее операции: включение элемента в дерево, поиск ключа в дереве, удаление ключа из дерева, отображение дерева, создание двоичного файла записей фиксированной длины из заранее подготовленных данных в текстовом файле, поиск записи в файле с использованием БДП.

Код приложения.

Код БДП:

```
compares = 0
 def get(self, key) -> int:
```

```
def get(self, key):
```

Класс обработки файла:

```
class FileHandler:
    def __init__ (self, file_name: str, data: Dict[str, int]):
        self.file_name = file_name
        self.data = data
        self.length = 0
        self.length = 0
        with open(self.file_name, 'w'):
            pass

    def add(self, **kwargs) -> str:
        res = []
        for key in self.data.keys():
            s = str(kwargs[key]).ljust(self.data[key], ' ')
            res.append(s)
        res = ''.join(map(str, res))
        with open(self.file_name, 'ab') as f:
            f.write(res.encode())
            self.length = len(res.encode())
        self.size += 1
        return self.size - 1

def get(self, n: int) -> str:
        pos = self.length * n
        with open(self.file_name, 'rb') as f:
            f.seek(pos)
            res = f.read(self.length).decode()
        return res

def remove(self, n: int) -> None:
        pass
```

Класс объединения работы с файлом и деревом:

```
class Combining:
    def __init__(self, combining, file_handler):
        self.combining = combining
        self.file_handler = file_handler

    def get(self, key) -> str:
        from time import perf_counter # Импорт perf_counter из библиотеки

time.
        a = perf_counter() # Время до получения записи.
        n = self.combining.get(key)
        b = perf_counter() # Время после получения записи.
        print(f"Потраченное время на получение записи: {b - a:0.7f} c.") #

Потраченное время на получение записи.
        if n != -1:
            return self.file_handler.get(n)
        else:
            return 'None'

def add(self, key, **kwargs) -> None:
        value = self.file_handler.add(key=key, **kwargs)
        self.combining.add(key, value)

def remove(self, key) -> None:
        value = self.combining.get(key)
        self.combining.remove(key)
        self.file handler.remove(value)
```

Код тестирования:

```
from BSTree import BSTree as Tree
from FileHandler import FileHandler
def test():
   f.close() # Закрытие файла "data.txt".
def fill comb(comb: 'Combining', path: str):
           line = line.split(';')
```

Тестирование программы.

Было запущено тестирование, в ходе которого в файл с данными было записано 10000 элементов (Рисунок 1).

```
1103033083; Place3399; Place6424; Time8535
1103030577; Place3313; Place7316; Time3930
1103038755; Place2056; Place5164; Time6807
1103036943; Place3915; Place9843; Time6671
1103030978; Place418; Place9020; Time1074
1103039628; Place3019; Place5204; Time5943
1103037550; Place522; Place6874; Time6578
1103038770; Place4648; Place9109; Time6513
1103037659; Place3304; Place5155; Time1287
1103031958; Place3721; Place6517; Time3788
1103030808; Place2509; Place5657; Time3756
1103037521; Place2360; Place8159; Time7313
1103039284; Place132; Place6885; Time3771
1103038746; Place3389; Place7949; Time3170
1103036162; Place2891; Place8806; Time9970
1103032890; Place1407; Place6091; Time5416
1103032037; Place104; Place7354; Time7266
1103030474; Place934; Place7025; Time1114
1103034230; Place2283; Place8521; Time3599
1103031248; Place3366; Place8192; Time7604
1103030431; Place2815; Place7987; Time5058
```

Рисунок 1 – Фрагмент файла из 10000 записей.

На основе файла было создано бинарное дерево поиска из 10000 узлов (Рисунок 2). Каждый узел дерева содержит в качестве ключа номер поезда, а в качестве информационного поля — номер строки в файле, который является ссылкой на информацию по ключу: пункт отправления, пункт назначения, время отправления.

```
1103032071|8002
1103032070|6836
1103032070|6836
1103032068|5041
1103032068|5041
1103032066|5611
1103032066|5611
1103032066|5611
1103032066|5631
1103032064|6537
1103032062|4126
1103032062|4126
1103032061|2127
1103032062|4126
1103032059|9478
1103032059|9478
1103032059|9478
1103032055|6088
1103032055|6088
1103032055|6088
1103032055|853
1103032052|8256
1103032052|8256
1103032052|8256
1103032052|8155
1103032049|811
1103032049|811
1103032049|811
1103032049|811
1103032049|811
1103032049|11
1103032049|115
1103032049|506
1103032049|506
1103032040|7126
1103032040|7176
1103032039|3342
1103032039|3509
1103032039|35509
```

Рисунок 2 – Фрагмент БДП из 10000 узлов.

Далее по бинарному дереву поиска были получены первая запись файла, последняя запись файла и запись, расположенная в середине файла. Для каждой полученной записи было найдено потраченное время на её получение, а также выведено количество произведённых сравнений во время получения записи (Рисунок 3).

```
Получение первой записи:
Потраченное время на получение записи: 0.0000044 с.
1103032245 Place4487 Place5473 Time8596

Число произведённых сравнений при получении первой записи: 1

Получение последней записи:
Потраченное время на получение записи: 0.0000140 с.
1103030431 Place2815 Place7987 Time5058

Число произведённых сравнений при получении последней записи: 16

Получение записи, расположенной в середине файла:
Потраченное время на получение записи: 0.0000101 с.
1103039876 Place3988 Place5282 Time484

Число произведённых сравнений при получении записи, расположенной в середине файла: 14
```

Рисунок 3 – Полученные данные, время поиска и количество сравнений.

Задание 2

Постановка задачи.

Разработать приложение, которое использует сбалансированное дерево поиска, предложенное в варианте, для доступа к записям файла.

Вариант 11.

Сбалансированное дерево	Структура элемента множества
поиска (СДП)	(ключ – подчеркнутое поле)
	Железнодорожная справка:
Красно-чёрное	номер поезда, пункт отправления,
	пункт назначения, время отправления

Дано:

Класс «Сбалансированное дерево поиска». Методы: включение элемента в дерево, поиск ключа в дереве, удаление ключа из дерева, отображение дерева.

Класс управления файлом. Методы: создание двоичного файла записей фиксированной длины из заранее подготовленных данных в текстовом файле; поиск записи в файле с использованием СДП.

Результат.

Приложение, выполняющее операции: включение элемента в дерево, поиск ключа в дереве, удаление ключа из дерева, отображение дерева, создание двоичного файла записей фиксированной длины из заранее подготовленных данных в текстовом файле, поиск записи в файле с использованием СДП, нахождение количества выполненных поворотов дерева.

Код приложения.

Код СДП (красно-чёрного дерева):

```
BLACK = 'ЧЁРНЫЙ'
    def paint(self, color):
    def isLeftChild(self):
    def sibling(self):
        return self.parent.sibling()
                self.left = RBNode(key, value, parent, color)
```

```
node.paint(BLACK)
    uncle = parent.sibling()
        if node.isLeftChild():
            node.paint(BLACK)
            self.RightRotate(parent)
def get(self, key) -> int:
```

```
sibling.right.paint(BLACK)
```

```
sibling = parent.left
def LeftRotate(self, grand):
```

Код тестирования:

```
from RBTree import RBTree as Tree
from FileHandler import FileHandler
def test():
    f.close() # Закрытие файла "data.txt".
   fw = FileHandler('data.bin', {'key': 11, 'place1': 10, 'place2': 10,
RBTree.compares | ')
def fill comb(comb: 'Combining', path: str):
           line = line.split(';')
```

Тестирование программы.

Было запущено тестирование, в ходе которого в файл с данными было записано 10000 элементов (Рисунок 4).

```
1103032167; Place1938; Place9441; Time846
1103037855; Place3723; Place5979; Time8234
1103033413; Place4769; Place7466; Time551
1103033057; Place2085; Place6895; Time602
1103033756; Place1270; Place9343; Time4438
1103036747; Place934; Place6983; Time8239
1103031460; Place3515; Place6591; Time5966
1103034655; Place4330; Place7233; Time7642
1103033612; Place4882; Place5818; Time6679
1103031557; Place3013; Place9227; Time3535
1103031363; Place1965; Place9853; Time8181
1103031881; Place2430; Place7615; Time1907
1103031831; Place3810; Place5813; Time6020
1103030275; Place4023; Place5181; Time6347
1103031823; Place452; Place8901; Time1569
1103032300; Place3932; Place8643; Time5177
1103031302; Place1014; Place5358; Time2925
1103031200; Place2136; Place8696; Time3962
1103031621; Place2543; Place5566; Time5639
1103038309; Place4794; Place7142; Time3420
1103039925;Place1925;Place8430;Time5704
```

Рисунок 4 – Фрагмент файла из 10000 записей.

На основе файла было создано СДП: красно-чёрное дерево (Рисунок 5). Каждый узел дерева содержит информацию о своём цвете (узел – красный или чёрный), ключ – номер поезда, а также информационное поле – номер строки в файле, который является ссылкой на информацию по ключу: пункт отправления, пункт назначения, время отправления.

```
-чёрный-1103030053|850
                                   -красный-1103030052|6183
                                   -красный-110303005015960
                           -чёрный-1103030049|5566
                                   -красный-1103030048|8722
         -чёрный-1103030047|3400
                                   -красный-1103030046|8388
                          -чёрный-1103030045|2179
                  -красный-1103030044|4604
                          -чёрный-1103030043|6656
-красный-1103030042|1035
                          -чёрный-1103030041|7908
                                   -красный-1103030040|8321
                          -чёрный-1103030038|2824
         -чёрный-1103030037|2137
                                    -красный-1103030036|4612
                                   -красный-1103030034|2643
                 -красный-1103030033|146
                                    -красный-1103030032|2595
                           -чёрный-1103030031|5169
                                -красный-1103030030|9631
```

Рисунок 5 – Фрагмент СДП, построенного по файлу из 10000 записей.

Далее по дереву были получены первая запись файла, последняя запись файла и запись, расположенная в середине файла. Для каждой полученной записи было найдено потраченное время на её получение, а также выведено количество произведённых сравнений во время получения записи (Рисунок 6).

```
Получение первой записи:
Потраченное время на получение записи: 0.0000079 с.
1103039288 Place2508 Place9938 Time2419

Число произведённых сравнений при получении первой записи: 5

Получение последней записи:
Потраченное время на получение записи: 0.0000111 с.
1103039925 Place1925 Place8430 Time5704

Число произведённых сравнений при получении последней записи: 14

Получение записи, расположенной в середине файла:
Потраченное время на получение записи: 0.0000101 с.
1103031445 Place3516 Place6636 Time3685

Число произведённых сравнений при получении записи, расположенной в середине файла: 13
```

Рисунок 6 – Полученные данные, время поиска и количество сравнений.

После этого было получено общее количество произведённых поворотов и выведено на экран (Рисунок 7).

```
Общее число произведённых поворотов: 5739
```

Рисунок 7 – Количество произведённых поворотов.

Задание 3

Постановка задачи.

Выполнить анализ алгоритма поиска записи с заданным ключом при применении структур данных:

- хеш-таблица;
- бинарное дерево поиска;
- СДП (красно-чёрное дерево).

Код приложения.

Код тестирования:

```
def test():
   f.close() # Закрытие файла "data.txt".
   fw = FileHandler('data.bin', {'key': 11, 'place1': 10, 'place2': 10,
```

```
ht = HT() # Создание хеш-таблицы.
bs_comb = Combining(bs, fw) # Объединение работы файла и БДП.
rb_comb = Combining(rb, fw) # Объединение работы файла и СДП.
ht_comb = Combining(ht, fw) # Объединение работы файла и хеш-таблицы.
```

```
инарного дерева поиска:")
print("
'time1': 9})
bs = BST() # Создание БДП.
     rb = RBT()  # Создание СДП.
ht = HT()  # Создание хеш-таблицы.
```

Тестирование программы.

Было запущено тестирование, в ходе которого в файл с данными было записано 1000 элементов (Рисунок 8).

```
110303097; Place111; Place892; Time792
110303234; Place417; Place957; Time283
110303261; Place123; Place638; Time7
110303314; Place421; Place803; Time33
110303882; Place440; Place675; Time623
110303959; Place257; Place902; Time737
110303645; Place138; Place926; Time897
110303996; Place416; Place535; Time108
110303699; Place108; Place947; Time745
110303268; Place460; Place775; Time433
110303803; Place65; Place625; Time946
110303935; Place195; Place514; Time841
110303695; Place275; Place818; Time248
110303753; Place314; Place784; Time585
110303794; Place167; Place963; Time574
110303979; Place413; Place877; Time281
110303056; Place145; Place639; Time865
110303194; Place11; Place853; Time341
110303328; Place438; Place869; Time853
110303426; Place188; Place907; Time249
```

Рисунок 8 – Фрагмент файла из 1000 записей.

На основе файла были созданы следующие структуры данных: бинарное дерево поиска, сбалансированное дерево поиска (красно-чёрное дерево) и хеш-таблица с цепным хешированием.

Далее была получена первая запись файла: сначала с помощью БДП, затем с помощью СДП и после этого с помощью хеш-таблицы. Для каждого алгоритма получения записи было найдено потраченное время на её получение, а также выведено количество произведённых сравнений во время получения записи (Рисунок 9).

```
Запущено тестирование программы для случая, когда в файле 1000 записей.

(000)

Получение первой записи с помощью бинарного дерева поиска:

Потраченное время на получение записи: 0.0000050 с.

110303142 Р1асе426 Р1асе979 Тіте388

Число произведённых сравнений при получении первой записи: 1

Получение первой записи с помощью СДП (красно-чёрного дерева):

Потраченное время на получение записи: 0.0000074 с.

110303142 Р1асе426 Р1асе979 Тіте388

Число произведённых сравнений при получении первой записи: 3

Получение первой записи с помощью жеш-таблицы:

Потраченное время на получение записи: 0.0004338 с.

110303142 Р1асе426 Р1асе979 Тіте388
```

Рисунок 9 – Полученные данные, время поиска и количество сравнений для первой записи.

Подобным образом были получены последняя запись и запись, расположенная в середине файла, а также соответствующая информация по работе алгоритмов поиска. (Рисунок 10, 11).

Рисунок 10 – Полученные данные, время поиска и количество сравнений для последней записи.

Рисунок 11 – Полученные данные, время поиска и количество сравнений для записи, расположенной в середине файла.

После этого был обновлён файл с данными: в него было записано теперь 10000 элементов (Рисунок 12).

```
1103032373; Place1368; Place8035; Time1896
1103037204; Place1761; Place8968; Time2660
1103031529; Place4103; Place8833; Time4187
1103034311;Place3485;Place9175;Time890
1103032763; Place3184; Place7805; Time3450
1103034229; Place307; Place8356; Time789
1103034914; Place4343; Place8656; Time6726
1103037060; Place2952; Place5407; Time2128
1103038295; Place129; Place5582; Time7828
1103033254; Place299; Place9704; Time7606
1103031536; Place495; Place5220; Time7921
1103033725; Place1296; Place7436; Time6593
1103038333; Place4109; Place5762; Time3768
1103030076; Place3791; Place9755; Time7144
1103039366; Place2181; Place7735; Time266
1103030536; Place2188; Place9230; Time7672
1103033686; Place2586; Place8863; Time6084
1103032383; Place4080; Place5080; Time6177
1103034816; Place2736; Place6142; Time4401
1103034570; Place1339; Place9461; Time4816
1103031538; Place4863; Place8638; Time4980
```

Рисунок 12 – Фрагмент файла из 10000 записей.

На основе файла были созданы структуры данных: бинарное дерево поиска, сбалансированное дерево поиска (красно-чёрное дерево) и хештаблица с цепным хешированием.

Далее была получена первая запись файла: сначала с помощью БДП, затем с помощью СДП и после этого с помощью хеш-таблицы. Для каждого алгоритма получения записи было найдено потраченное время на её получение, а также выведено количество произведённых сравнений во время получения записи (Рисунок 13).

Рисунок 13 — Полученные данные, время поиска и количество сравнений для первой записи.

Подобным образом были получены последняя запись и запись, расположенная в середине файла, а также соответствующая информация по работе алгоритмов поиска. (Рисунок 14, 15).

```
Получение последней записи с помощью бинарного дерева поиска:

Потраченное время на получение записи: 0.0000296 с.
1103031538 Place4863 Place8638 Time4980

Число произведённых сравнений при получении последней записи: 21

Получение последней записи с помощью СДП (красно-чёрного дерева):

Потраченное время на получение записи: 0.0000126 с.
1103031538 Place4863 Place8638 Time4980

Число произведённых сравнений при получении последней записи: 15

Получение последней записи с помощью хеш-таблицы:

Потраченное время на получение записи: 0.0011585 с.
1103031538 Place4863 Place8638 Time4980

Число произведённых сравнений при получении последней записи: 2
```

Рисунок 14 — Полученные данные, время поиска и количество сравнений для последней записи.

Рисунок 15 — Полученные данные, время поиска и количество сравнений для записи, расположенной в середине файла.

Анализ результатов.

Получение первой записи из файла:

Количество				
элементов,	Вид поисковой структуры	Емкостная		
загруженных		сложность:	Количество	Время на
в структуру		объем	выполненных	поиск ключа в
в момент		памяти для	сравнений	структуре, мс
выполнения		структуры		
поиска				
	БДП	n	1	0.0050
1000	Красно-чёрное	n	3	0.0074
	Хэш-таблица	n	32	0.4338
	БДП	n	1	0.0065
10000	Красно-чёрное	n	1	0.0048
	Хэш-таблица	n	156	1.4702

Получение последней записи из файла:

Количество				
элементов,	Вид поисковой структуры	Емкостная		
загруженных		сложность:	Количество	Время на
в структуру		объем	выполненных	поиск ключа в
в момент		памяти для	сравнений	структуре, мс
выполнения		структуры		
поиска				
	БДП	n	16	0.0190
1000	Красно-чёрное	n	11	0.0137
	Хэш-таблица	n	2	0.3646
	БДП	n	21	0.0296
10000	Красно-чёрное	n	15	0.0126
	Хэш-таблица	n	2	1.1585

Получение записи, расположенной в середине файла:

Количество				
элементов,	Вид поисковой структуры	Емкостная		
загруженных		сложность:	Количество	Время на
в структуру		объем	выполненных	поиск ключа в
в момент		памяти для	сравнений	структуре, мс
выполнения		структуры		
поиска				
	БДП	n	11	0.0141
1000	Красно-чёрное	n	10	0.0120
	Хэш-таблица	n	18	0.4093
	БДП	n	13	0.0117
10000	Красно-чёрное	n	15	0.0117
	Хэш-таблица	n	82	1.1611

По результатам видно, что почти во всех случаях наиболее быстрое получение данных осуществлялось с помощью сбалансированного дерева поиска: красно-чёрного дерева. В общем случае, чуть менее быстрым был поиск данных посредством бинарного дерева поиска. Получение данных с помощью хеш-таблицы происходило медленнее во всех случаях.

Вывод

В ходе работы были приобретены практические навыки по использованию и построению бинарного дерева поиска и сбалансированного дерева поиска.

Список информационных источников

- 1. Лекции по дисциплине «Структуры и алгоритмы обработки данных» /
- Л. А. Скворцова, МИРЭА Российский технологический университет, 2021.