Отчет о выполнении задания 2  
«Бинарные деревья поиска и хеш-таблицы»

**Чуклин Алексей Сергеевич**

Группа ИВ-521

cas9833@gmail.com

# Описание алгоритмов

В задании 2 требовалось реализовать и исследовать эффективность бинарных деревьев поиска и хеш-таблиц. Все алгоритмы реализованы на языке C под операционной системой GNU/Linux.

# Бинарное дерево поиска[1]

Бинарное дерево поиска ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA" \o "Английский язык) binary search tree, BST) — это [двоичное дерево](https://ru.wikipedia.org/wiki/%D0%94%D0%B2%D0%BE%D0%B8%D1%87%D0%BD%D0%BE%D0%B5_%D0%B4%D0%B5%D1%80%D0%B5%D0%B2%D0%BE" \o "Двоичное дерево), для которого выполняются следующие дополнительные условия (свойства дерева поиска):

* Оба поддерева — левое и правое — являются двоичными деревьями поиска.
* У всех узлов левого поддерева произвольного узла X значения ключей данных меньше, нежели значение ключа данных самого узла X.
* В то время, как значения ключей данных у всех узлов правого поддерева (того же узла X) больше, нежели значение ключа данных узла X.

В работе реализована версия алгоритма с вычислительной сложностью следующих операций:

* Создание элемента дерева (bstree\_create): *O*(1);
* Добавление элемента (bstree\_add): *O*(*n*);
* Поиск элемента (bstree\_lookup): *O*(*n*);
* Поиск минимального значения (bstree\_min): *O*(*n*);
* Поиск максимального значения (bstree\_max): *O*(*n*);

Расход по памяти *O*(*n*), где *n* — это количество внесенных элементов в дерево.

# Хеш-таблицы[2]

Хеш-таблица — это [структура данных](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D0%B0_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85" \o "Структура данных), реализующая интерфейс [ассоциативного массива](https://ru.wikipedia.org/wiki/%D0%90%D1%81%D1%81%D0%BE%D1%86%D0%B8%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D1%8B%D0%B9_%D0%BC%D0%B0%D1%81%D1%81%D0%B8%D0%B2" \o "Ассоциативный массив), а именно, она позволяет хранить пары (ключ, значение) и выполнять три операции: операцию добавления новой пары, операцию поиска и операцию удаления пары по ключу.

В работе реализована версия алгоритма с вычислительной сложностью следующих операций:

* Создание хеша (hashtab\_hash): *O*(s);
* Добавление элемента (hashtab\_add): *O*(1);
* Поиск элемента (hashtab\_lookup): *O*(*n*);
* Удаление элемента: *O*(*n*);

Расход по памяти *O*(*n*), где *n* — это количество внесенных элементов в таблицу.

# Организация экспериментов

* Эксперименты проводились на ноутбуке  
  (CPU: i5-4210U @ 1.7GHz, RAM: 4GB)
* Операционная система kubuntu 16.04 (компилятор gcc 5.3.1)
* Ключи компиляции программы: -o

# Результаты экспериментов

Результаты экспериментов приведены в таблицах 1, 2 и 3 и в графиках 1, 2, 3.

*Таблица 1. Результаты поиска элемента и бинарном дереве и в хеш-таблице*

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Количество элементов** | **Бинарное дерево** | **Хеш-таблица** |
| 1 | 10000 | 0,000125 | 0,0000025 |
| 2 | 20000 | 0,000148 | 0,0000042 |
| 3 | 30000 | 0,000266 | 0,0000038 |
| 4 | 40000 | 0,000544 | 0,0000082 |
| 5 | 50000 | 0,000406 | 0,0000071 |
| 6 | 60000 | 0,000546 | 0,0000121 |
| 7 | 70000 | 0,000385 | 0,0000062 |
| 8 | 80000 | 0,000788 | 0,0000099 |
| 9 | 90000 | 0,000728 | 0,0000092 |
| 10 | 100000 | 0,000828 | 0,0000124 |
| 11 | 110000 | 0,000859 | 0,0000125 |
| 12 | 120000 | 0,001054 | 0,0000318 |
| 13 | 130000 | 0,001282 | 0,0000134 |
| 14 | 140000 | 0,00114 | 0,0000192 |
| 15 | 150000 | 0,001159 | 0,0000328 |
| 16 | 160000 | 0,001127 | 0,0000221 |
| 17 | 170000 | 0,001185 | 0,0000402 |
| 18 | 180000 | 0,001671 | 0,0000233 |
| 19 | 190000 | 0,001692 | 0,0000152 |
| 20 | 200000 | 0,001866 | 0,0000260 |

В данной таблице мы можем увидеть, что поиск в бинарном дереве занимает больше времени, относительно поиск хеш-таблицах.

*График 1. Результаты поиска элемента и бинарном дереве и в хеш-таблице*

*Таблица 2. Время добавления элементов в бинарном дереве в худшем и в среднем случаях*

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Количество элементов** | **Время выполнения функции bstree\_add, c** | **Время выполнения функции hashtab\_add, c** |
| 1 | 10000 | 0,241118 | 0,001869 |
| 2 | 20000 | 0,977776 | 0,003531 |
| 3 | 30000 | *2,21781* | 0,004751 |
| 4 | 40000 | *3,931897* | 0,005959 |
| 5 | 50000 | *6,160801* | 0,007147 |
| 6 | 60000 | *8,943717* | 0,008315 |
| 7 | 70000 | *12,316581* | 0,009484 |
| 8 | 80000 | *16,306091* | 0,010682 |
| 9 | 90000 | *20,950658* | 0,011848 |
| 10 | 100000 | *26,223729* | 0,013018 |
| 11 | 110000 | *32,127240* | 0,014184 |
| 12 | 120000 | *38,650013* | 0,015412 |
| 13 | 130000 | *46,686457* | 0,016602 |
| 14 | 140000 | *54,961276* | 0,017792 |
| 15 | 150000 | *63,388546* | 0,018969 |
| 16 | 160000 | *72,250999* | 0,020150 |
| 17 | 170000 | *81,727934* | 0,021317 |
| 18 | 180000 | *91,757355* | 0,022504 |
| 19 | 190000 | *102,382918* | 0,023680 |
| 20 | 200000 | *113,597959* | 0,024889 |

Из таблицы 2 мы можем увидеть, что добавление элемента в бинарном дереве занимает больше времени, чем в хеш-таблицах.

*График 2. Время поиска в бинарном дереве в худшем и в среднем случаях*

*Таблица 3. Время поиска в хеш-таблицах в алгоритмах KP и XOR*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Количество элементов** | **Хеш-функция KP** | | **Хеш-функция XOR** | |
| Время поиска | Коллизии | Время поиска | Коллизии |
| 1 | 10000 | 0,000002 | 8 | 0,000000 | 3 |
| 2 | 20000 | 0,000001 | 20 | 0,000000 | 9 |
| 3 | 30000 | 0,000001 | 31 | 0,000164 | 11 |
| 4 | 40000 | 0,000003 | 39 | 0,000256 | 13 |
| 5 | 50000 | 0,000004 | 48 | 0,000001 | 16 |
| 6 | 60000 | 0,000004 | 61 | 0,000418 | 19 |
| 7 | 70000 | 0,000016 | 68 | 0,000001 | 22 |
| 8 | 80000 | 0,000016 | 78 | 0,000581 | 26 |
| 9 | 90000 | 0,000017 | 86 | 0,000642 | 31 |
| 10 | 100000 | 0,000008 | 99 | 0,000786 | 31 |
| 11 | 110000 | 0,000012 | 106 | 0,000902 | 34 |
| 12 | 120000 | 0,000023 | 115 | 0,001383 | 38 |
| 13 | 130000 | 0,000024 | 129 | 0,000959 | 42 |
| 14 | 140000 | 0,000014 | 140 | 0,001302 | 45 |
| 15 | 150000 | 0,000024 | 152 | 0,0001124 | 47 |
| 16 | 160000 | 0,000014 | 161 | 0,0001271 | 52 |
| 17 | 170000 | 0,000026 | 170 | 0,0000008 | 55 |
| 18 | 180000 | 0,000025 | 188 | 0,0001283 | 59 |
| 19 | 190000 | 0,000024 | 193 | 0,0001299 | 61 |
| 20 | 200000 | 0,000017 | 198 | 0,0001399 | 62 |

По таблице видно, что в алгоритме хеш-функции DJB поиск медленнее, чем в хеш-таблице, построенной по алгоритму KP, но коллизий там меньше.

*График 3. Время поиска в хеш-таблицах в алгоритмах KP и XOR*

# Вывод

1. Поиск в бинарном дереве занимает больше времени, относительно поиска в хеш-таблицах.
2. Добавление элемента в бинарном дереве занимает больше времени, чем в хеш-таблицах.
3. В алгоритме хеш-функции DJB поиск медленнее, чем в хеш-таблице, построенной по алгоритму KP, но коллизий там меньше.

# Ссылки

1. <https://ru.wikipedia.org/wiki/Двоичное_дерево_поиска>
2. <https://ru.wikipedia.org/wiki/Хеш-таблица>