

Дослідження Self-Balancing Binary Search Trees

Кафедра комп'ютерних наук
Український католицький університет
Львів, Україна

Зміст

I	Вступ	3
II	Склад команди й розподіл задач	3
III	Мета та завдання	3
III-A	Мета	3
III-B	Основні задачі	3
IV	B-дерево	3
IV-A	Структура даних	3
IV-B	Опис реалізації функції BTree	3
V	Червно-чорне дерево	5
V-A	Теоретичні основи	5
V-B	Опис реалізації функцій RedBlackTree	5
VI	AVL-дерево	7
VI-A	Теоретичні основи	7
VI-B	Структура даних	7
VI-B	Опис реалізації	7
VI-Г	Переваги	7
VI-Д	Недоліки	7
VII	Splay-дерево	8
VIII	Обходи дерев	8
VIII-A	Прямий обхід (Preorder)	8
VIII-B	Симетричний обхід (Inorder)	8
VIII-B	Обернений обхід (Postorder)	8
VIII-Г	Порівняння	8
IX	Візуалізація продуктивності дерев	9
IX-A	Методологія тестування	9
IX-B	Результати візуалізації	9
IX-B	Інтерпретація результатів	9
IX-Г	Висновки	9
X	Розробка консольного інтерфейсу	10
X-A	Архітектура CLI	10
X-B	Підтримувані операції	10
X-B	Приклади використання	10
X-Г	Реалізація парсера запитів	10
X-Д	Вибір структури дерева	10
X-E	Тестовий режим	10

Анотація—У цьому звіті наведено опис та реалізацію чотирьох видів самобалансуювальних дерев пошуку: В-дерево, AVL-дерево, Червоно-чорне дерево та Splay-дерево.

I. Вступ

У даному проєкті здійснено дослідження самобалансованих бінарних дерев пошуку — фундаментальних структур даних, що гарантують $\mathcal{O}(\log n)$ -час на операції пошуку, вставки, видалення й обходу. Окрім теоретичного аналізу, реалізовано примітивну SQL-подібну систему бази даних, в якій індексація та зберігання даних виконуються безпосередньо на основі власних реалізацій дерев.

II. Склад команди й розподіл задач

- Падучак Маргарита — В-дерево
- Шевчук Дарина — AVL-дерево, візуалізація
- Ягода Микита — Red-Black Tree, розробка бази даних
- Лещук Роман — Splay-дерево, 2-3-дерево, тестування, допомога з базою даних

III. Мета та завдання

A. Мета

- 1) Вивчити принципи самобалансування дерев пошуку.
- 2) Розробити й оптимізувати п'ять різновидів таких дерев.
- 3) Створити на їхній основі просту СУБД із SQL-інтерфейсом.
- 4) Провести експерименти й порівняльний аналіз ефективності.

Б. Основні задачі

- Реалізувати структури: AVL-дерево, В-дерево, Red-Black Tree, Splay-дерево, 2-3-дерево.
- Забезпечити в кожній структурі: вставку, видалення, пошук, обходи (in-order, pre-order).
- Розробити консольний інтерфейс із командами INSERT, DELETE, SELECT, UPDATE.
- Зібрати статистику часу виконання операцій на різних обсягах даних.
- Порівняти показники з відкритими СУБД (MySQL, PostgreSQL).
- Побудувати графіки та таблиці для наочності результатів.

IV. В-дерево

A. Структура даних

В-дерево — це m -арне дерево пошуку, у якому:

- Кожен вузол (окрім кореня) має щонайменше $\lceil m/2 \rceil - 1$ та щонайбільше $m - 1$ ключів.
- Внутрішній вузол з k ключами має $k + 1$ дітей.
- Всі листи знаходяться на одній глибині.

Завдяки цьому В-дерево мінімізує кількість доступів до зовнішньої пам'яті і забезпечує $\mathcal{O}(\log n)$ час операцій.

- BTreeNode — клас, що представляє вузол В-дерева.
 - leaf (bool) — булевий прапорець, що вказує, чи є вузол листом.
 - keys (list) — список ключів, які зберігаються у вузлі. Ключі зберігаються у відсортованому порядку.
 - children (list) — список посилань на нащадків вузла (інші об'єкти BTreeNode).
- BTree — клас, що представляє саме В-дерево.
 - root — кореневий вузол дерева, об'єкт класу BTreeNode.
 - t — мінімальний ступінь дерева: кожен вузол (крім кореня) має принаймні $t - 1$ ключів і не більше $2t - 1$ ключів. Обчислюється як $t = \lceil \frac{m}{2} \rceil$, де m — максимально допустима кількість дітей у вузлі.

Б. Опис реалізації функції BTree

1) Пошук: Функція виконує рекурсивний пошук ключа k у вузлі x В-дерева. Вона переглядає ключі у вузлі зліва направо, доки не знайде ключ або не визначить, що потрібно перейти до відповідного нащадка. Якщо вузол є листком і ключ не знайдено — повертається None.

Algorithm 1 BTree_Find(k, x)

```

1: if  $x = \text{None}$  then
2:    $x \leftarrow \text{root}$ 
3: end if
4:  $i \leftarrow 0$ 
5: while  $i < |x.keys|$  and  $k >$ 
    $x.keys[i][0].columns[key\_col]$  do
6:    $i \leftarrow i + 1$ 
7: end while
8: if  $i < |x.keys|$  and  $k = x.keys[i][0].columns[key\_col]$ 
   then
9:   return  $i$ 
10: else if  $x.leaf$  then
11:   return None
12: else
13:   return BTree_Find( $k, x.children[i]$ )
14: end if

```

2) Вставка: Вставка в В-дерево починається з пошуку позиції для нового елемента. Якщо ключ уже існує, нове значення додається до списку відповідного вузла. Якщо корінь повний, він розділяється, і дерево збільшується вгору. Далі вставка виконується у відповідний підвузол, гарантуючи, що новий ключ завжди потрапляє в неповний вузол.

Algorithm 2 BTree_Insert(k)

```

1: existing  $\leftarrow$  BTree_Find_For_Insert( $k$ )
2: if existing  $\neq$  None then
3:   existing[0].keys[existing[1]].append( $k$ )
4:   return
5: end if
6: if |root.keys| =  $2t - 1$  then
7:   temp  $\leftarrow$  new BTreeNode(leaf = False)
8:   prev_root  $\leftarrow$  root
9:   root  $\leftarrow$  temp
10:  temp.children  $\leftarrow$  [prev_root]
11:  Split_Children(temp, 0)
12:  Insert_Non_Full(temp,  $k$ )
13: else
14:   Insert_Non_Full(root,  $k$ )
15: end if

```

Algorithm 3 Insert_Non_Full(x, k)

```

1:  $i \leftarrow |x.keys| - 1$ 
2: if  $x.leaf$  then
3:   append None до  $x.keys$ 
4:   while  $i \geq 0$  and  $k.columns[key\_col] <$ 
      $x.keys[i][0].columns[key\_col]$  do
5:      $x.keys[i+1] \leftarrow x.keys[i]$ 
6:      $i \leftarrow i - 1$ 
7:   end while
8:    $x.keys[i+1] \leftarrow [k]$ 
9: else
10:  while  $i \geq 0$  and  $k.columns[key\_col] <$ 
     $x.keys[i][0].columns[key\_col]$  do
11:     $i \leftarrow i - 1$ 
12:  end while
13:   $i \leftarrow i + 1$ 
14:  if  $|x.children[i].keys| = 2t - 1$  then
15:    Split_Children( $x, i$ )
16:    if  $k.columns[key\_col] >$ 
        $x.keys[i][0].columns[key\_col]$  then
17:       $i \leftarrow i + 1$ 
18:    end if
19:  end if
20:  Insert_Non_Full( $x.children[i], k$ )
21: end if

```

3) Розбиття вузла: Розбиття (розщеплення) вузла в В-дереві виконується тоді, коли вузол переповнений. Його середній ключ піднімається в батьківський вузол, а сам вузол розділяється на два: лівий з ключами до

середнього та правий з ключами після. Якщо це не лист, відповідно перерозподіляються і нащадки.

Algorithm 4 Split_Children(x, i)

```

1:  $t \leftarrow \text{self.t}$ 
2:  $y \leftarrow x.children[i]$ 
3:  $z \leftarrow$  new BTreeNode(leaf =  $y.leaf$ )
4: insert  $z$  у  $x.children$  на позицію  $i + 1$ 
5: insert  $y.keys[t - 1]$  у  $x.keys$  на позицію  $i$ 
6:  $z.keys \leftarrow y.keys[t : 2t]$ 
7:  $y.keys \leftarrow y.keys[0 : t - 1]$ 
8: if  $\neg y.leaf$  then
9:    $z.children \leftarrow y.children[t : 2t]$ 
10:   $y.children \leftarrow y.children[0 : t]$ 
11: end if

```

4) Видалення: Видалення елемента з В-дерева виконується так, щоб зберегти баланс і властивості дерева. Якщо ключ знаходиться в листі — його просто видаляють. Якщо в внутрішньому вузлі — замінюють на попередник або наступник і рекурсивно видаляють. Якщо дочірній вузол має мінімальну кількість ключів, виконується злиття або позичання ключа у сусіда для підтримання мінімального розміру.

Algorithm 5 BTree_Delete(x, k)

```
1:  $i \leftarrow 0$ 
2: while  $i < |x.keys|$  and  $k > x.keys[i]$  do
3:    $i \leftarrow i + 1$ 
4: end while
5: if  $i < |x.keys|$  and  $k = x.keys[i]$  then
6:   if  $x.leaf$  then
7:     Видалити  $k$  з  $x.keys$ 
8:   else
9:     if  $|x.children[i].keys| \geq t$  then
10:       $k' \leftarrow$  попередник  $k$ 
11:      BTree_Delete( $x.children[i], k'$ )
12:       $x.keys[i] \leftarrow k'$ 
13:     else if  $|x.children[i + 1].keys| \geq t$  then
14:       $k' \leftarrow$  наступник  $k$ 
15:      BTree_Delete( $x.children[i + 1], k'$ )
16:       $x.keys[i] \leftarrow k'$ 
17:     else
18:       Об'єднати  $x.children[i]$ ,  $k$  та  $x.children[i + 1]$ 
19:       BTree_Delete( $x.children[i], k$ )
20:     end if
21:   end if
22: else
23:   if  $x.leaf$  then
24:     Повернути — ключ не знайдено
25:   else
26:     if  $|x.children[i].keys| = t - 1$  then
27:       if  $i > 0$  and  $|x.children[i - 1].keys| \geq t$  then
28:         Позичити ключ з  $x.children[i - 1]$ 
29:       else if  $i < |x.children| - 1$  and  $|x.children[i + 1].keys| \geq t$  then
30:         Позичити ключ з  $x.children[i + 1]$ 
31:       else
32:         Об'єднати  $x.children[i]$  з сусіднім вузлом
33:       end if
34:     end if
35:     BTree_Delete( $x.children[i], k$ )
36:   end if
37: end if
```

V. Червно-чорне дерево**A. Теоретичні основи**

Червоно-чорне дерево – різновид самозбалансованого бінарного дерева пошуку, вершини якого мають додаткову властивість – колір. В червоно-чорному дереві також є додаткові правила:

- 1) Кожен вузол або чорний, або червоний
- 2) Корінь – чорний
- 3) Кожний ніл листок – чорний
- 4) Якщо вершина червона – то обидва її сини чорні
- 5) Усі прості шляхи від будь-якої вершини до будь-якого листка в її лівому та правому піддереві містять однакову кількість чорних вершин, не враховуючи саму вершину.

B. Опис реалізації функцій RedBlackTree

- 1) Вставка: Інтуїція та загальна ідея

Вставляємо новий червоний вузол як в бінарному дереві пошуку. Якщо батько нової вершини чорного кольору – жодних правил червоно-чорного дерева не порушено, а отже вставка закінчена. Інакше – порушено правило, що сини червоного вузла чорні – виправляємо порушення в піддеревах, поки загальне дерево не виконуватиме всіх правил. В кожен момент може бути порушення лише одного правила.

Algorithm 6 Insert(x)

```
1: Вставити новий Node червоного кольору як в бінарному дереві пошуку
2: if  $parent.color == BLACK$  then
3:   return
4: else
5:   if  $uncle.color == RED$  then
6:      $uncle.color = parent.color = BLACK$ 
7:      $grandparent.color = RED$ 
8:     рекурсивно виправляємо можливі порушення правил з вузла grandparent
9:   else
10:    if parent is left child and node is right child then
11:      Rotate parent left
12:      рекурсивно виправляємо порушення правил з вузла parent(який був до повороту)
13:    end if
14:    if parent is left child and node is left child then
15:      Rotate grandparent right
16:      рекурсивно виправляємо порушення правил з вузла node(який і був)
17:    end if
18:    Аналогічні перевірки та дії з іншого боку
19:  end if
20: end if
```

- 2) Видалення: Інтуїція та загальна ідея

Видаляємо як в бінарному дереві пошуку: якщо вершина має синів, то беремо значення(колір незмінний)

найменшого з правого піддерева, або найбільшого з лівого і тепер видаляємо того, чиє значення брали. Повторюємо до тих пір, поки не маємо видалити листок. Якщо листок червоний, то нічого не робимо, оскільки жодне правило не порушене. Якщо чорний, то позначаємо цю вершину як Двічічорного(Doubleblack | DB) і після балансування видаляємо. Виправляємо до тих пір, поки DB не зникне, або не стане коренем.

Case	Check condition	Action
1	If node to be delete is a red leaf node	Just remove it from the tree
2	If DB node is root	Remove the DB and root node becomes black.
3	(a) If DB's sibling is black, and (b) DB's sibling's children are black	(a) Remove the DB (if null DB then delete the node and for other nodes remove the DB sign) (b) Make DB's sibling red . (c) If DB's parent is black, make it DB, else make it black
4	If DB's sibling is red	(a) Swap color DB's parent with DB's sibling (b) Perform rotation at parent node in the direction of DB node (c) Check which case can be applied to this new tree and perform that action
5	(a) DB's sibling is black (b) DB's sibling's child which is far from DB is black (c) DB's sibling's child which is near to DB is red	(a) Swap color of sibling with sibling's red child (b) Perform rotation at sibling node in direction opposite of DB node (c) Apply case 6
6	(a) DB's sibling is black, and (b) DB's sibling's far child is red (remember this node)	(a) Swap color of DB's parent with DB's sibling's color (b) Perform rotation at DB's parent in direction of DB (c) Remove DB sign and make the node normal black node (d) Change colour of DB's sibling's far red child to black.

Табл. 1
Red-Black Tree Deletion Cases

VI. AVL-дерево

А. Теоретичні основи

AVL-дерево — це перше самобалансувальне двійкове дерево пошуку, запропоноване Георгієм Адельсоном-Вельським і Євгеном Ландісом у 1962 році. Його основною властивістю є збереження балансу: для кожного вузла висота лівого і правого піддерева відрізняється не більше ніж на 1.

Кожна операція (вставка, видалення) зберігає цю властивість через відповідне балансування — обертання вузлів. Усі основні операції виконуються за час $O(\log n)$ у найгіршому випадку.

Б. Структура даних

- Кожен вузол зберігає об'єкт типу `DataEntry`.
- Містить вказівники на лівого та правого нащадка.
- Додаткове поле `height` визначає висоту вузла, необхідну для обчислення коефіцієнта балансу.

В. Опис реалізації

Клас `AVLTree` реалізує базовий інтерфейс `AbstractTree` та містить основні методи:

- `insert(entry)` — вставка елемента з балансуванням.
- `erase(key)` — видалення елементів за ключем.
- `find(key)` — пошук усіх відповідників.
- `inorder()`, `preorder()`, `postorder()` — різні обходи дерева.

1) Обчислення висоти та балансу:

```
def get_height(self, node):
    if not node:
        return 0
    return node.height
```

```
def get_balance(self, node):
    if not node:
        return 0
    return self.get_height(node.left) - self.get_height(node.right)
```

2) Обертання вузлів:

```
def __rotate_left(self, node):
    new_root = node.right
    node.right = new_root.left
    new_root.left = node
    update_heights...
    return new_root
```

```
def __rotate_right(self, node):
    new_root = node.left
    node.left = new_root.right
    new_root.right = node
    update_heights...
    return new_root
```

3) Вставка елемента:

```
def __insert(self, node, data_entry):
    if node is None:
        return new AVLTreeNode(data_entry)

    if key < node_key:
        node.left = insert(node.left, data_entry)
    elif key > node_key:
        node.right = insert(node.right, data_entry)
    else:
        node.data.append(data_entry)
        return node
```

```
update height and balance
perform rotations if needed
return node
```

4) Видалення елемента: Видалення враховує всі стандартні випадки: вузол без дітей, з одним або двома, з наступним балансуванням.

```
def __delete(self, node, key):
    if node is None:
        return None

    # locate and delete node
    # find successor if necessary
    # update height and balance
    # perform rotations
    return node
```

Г. Переваги

- Гарантована логарифмічна глибина.
- Підходить для частих вставок/видалень.

Д. Недоліки

- Ускладнена реалізація.
- Більше обчислень при вставці та видаленні в порівнянні з незбалансованими деревами.

VII. Splay-дерево

VIII. Обходи дерев

У деревоподібних структурах даних існує три класичні способи обходу: прямий (preorder), симетричний (inorder) та обернений (postorder). Ці обходи важливі для аналізу структури, копіювання, серіалізації та візуалізації дерев.

A. Прямий обхід (Preorder)

Порядок відвідування: корінь \rightarrow ліве піддерево \rightarrow праве піддерево.

```
def preorder(self) -> list[DataEntry]:
    def __preorder_recursive(node, result):
        if node is None:
            return
        result.extend(node.data)
        __preorder_recursive(node.left, result)
        __preorder_recursive(node.right, result)
    ...
```

Приклад використання: копіювання структури дерева або побудова виразів.

Б. Симетричний обхід (Inorder)

Порядок відвідування: ліве піддерево \rightarrow корінь \rightarrow праве піддерево.

```
def inorder(self) -> list[DataEntry]:
    def __inorder_recursive(node, result):
        if node is None:
            return
        __inorder_recursive(node.left, result)
        result.extend(node.data)
        __inorder_recursive(node.right, result)
    ...
```

Приклад використання: виведення ключів дерева у відсортованому порядку (у випадку двійкового дерева пошуку).

В. Обернений обхід (Postorder)

Порядок відвідування: ліве піддерево \rightarrow праве піддерево \rightarrow корінь.

```
def postorder(self) -> list[DataEntry]:
    def __postorder_recursive(node, result):
        if node is None:
            return
        __postorder_recursive(node.left, result)
        __postorder_recursive(node.right, result)
        result.extend(node.data)
    ...
```

Приклад використання: звільнення пам'яті або видалення дерева.

Г. Порівняння

Усі три види обходу реалізуються рекурсивно та мають складність $O(n)$, де n — кількість вузлів у дереві.

Обхід	Порядок
Preorder	Корінь, Ліво, Право
Inorder	Ліво, Корінь, Право
Postorder	Ліво, Право, Корінь

Табл. II

Порівняння алгоритмів обходу дерева

IX. Візуалізація продуктивності дерев

Для оцінки ефективності реалізованих дерев пошуку було розроблено скрипт тестування (visualization.py), який вимірює час виконання ключових операцій: вставка (insert), пошук (find) та видалення (erase) для різних розмірів даних.

A. Методологія тестування

Тестування проводилося на випадково згенерованих наборах даних, що складаються з об'єктів типу DataEntry. Для кожного з розмірів (від 100 до 10 000 записів) виконувались:

- серія вставок,
- серія пошуків випадкових ключів,
- серія видалень випадкових елементів (50% записів).

Кожен експеримент повторювався кілька разів для усереднення результатів.

B. Результати візуалізації

Скрипт генерує три типи графіків:

1) Звичайні графіки: Відображають залежність часу виконання від кількості елементів:

- Insert Performance — загальний час на вставку всіх елементів.
- Find Performance — середній час на одну операцію пошуку.
- Erase Performance — середній час на одну операцію видалення.

2) Графіки з логарифмічною шкалою: Ці графіки (log-log) дозволяють краще побачити різницю в асимптотичних властивостях між реалізаціями, особливо для великих розмірів.

3) Стовпчикові діаграми: Для порівняння продуктивності різних дерев на фіксованому розмірі (наприклад, $n = 10\,000$), скрипт будує стовпчикові діаграми, де кожна група включає:

- час вставки,
- час пошуку,
- час видалення.

B. Інтерпретація результатів

- AVL-дерево демонструє стабільно хорошу продуктивність на всіх операціях завдяки суворому балансуванню.
- Splay-дерево швидке при повторному доступі до тих самих елементів, але може бути повільніше на випадкових наборах.

- Червоні-чорне дерево трохи швидше на вставках і видаленнях у середньому, однак його ефективність сильно залежить від реалізації.

Г. Висновки

Візуалізація підтвердила теоретичні оцінки складності алгоритмів. Збалансовані дерева пошуку (AVL, Red-Black) забезпечують передбачувану та ефективну роботу, в той час як Splay-дерево краще підходить для сценаріїв з локальністю доступу.

Х. Розробка консольного інтерфейсу

А. Архітектура CLI

Модуль `crud.py` реалізує консольний інтерфейс командного рядка (CLI) для взаємодії з базою даних, що базується на самобалансуючих деревах. Він надає можливість виконувати SQL-подібні запити до даних, які зберігаються в структурах різних типів дерев.

1) Основні компоненти:

- Парсер запитів — інтерпретує SQL-подібні команди користувача.
- Модуль валідації — перевіряє коректність імен таблиць, стовпців та значень.
- Система обробки помилок — забезпечує інформативні повідомлення про помилки.
- Інтерактивний режим — дозволяє користувачу працювати через командний рядок.

Б. Підтримувані операції

Реалізовано базові SQL-операції, зокрема:

- `SELECT` — вибірка даних з колекцій за заданими критеріями.
- `INSERT` — додавання нових записів до таблиці.

В. Приклади використання

1) Вибірка даних:

```
python crud.py -t
python crud.py test_database rb -
q select id name from another_table
```

2) Додавання запису:

```
python crud.py -t
python crud.py test_database avl -
q insert into another_table values 5 name description
```

3) Інтерактивний режим:

```
python crud.py -t
python crud.py test_database splay -i
Query> select * from another_table
Query> insert into another_table values 10 new_item "sample description"
```

Г. Реалізація парсера запитів

Функція `parse_query` аналізує аргументи командного рядка та перетворює їх у відповідні операції бази даних

Д. Вибір структури дерева

CLI надає можливість вибирати тип дерева для використання в базі даних через параметр командного рядка:

```
TreeType = get_tree_class(tree_name)
```

Підтримуються всі реалізовані типи дерев:

- AVL
- B-Tree
- Red-Black
- Splay
- Treap

Е. Тестовий режим

Для спрощення тестування та демонстрації функціональності, модуль має вбудований тестовий режим, який створює приклад бази даних з тестовими таблицями та даними:

```
python crud.py -t
```

Модуль забезпечує надійну обробку помилок, що покращує досвід користувача при роботі з системою.