Datastructures and Algorithms

Binary Trees, Heaps, Hashing

Adel Gavranović — ETH Zürich — 2025

Overview

Learning Objectives
Binary Trees and Heaps
Hashing
Binary Tree: Simple Tasks
Code-Example: Hashtables, Hash-functions and Collisions
Past Exam Questions
Tips for **code** expert

n.ethz.ch/~agavranovic

Material

Webpage

Mail

1. Follow-up

Follow-up from last session

- Regarding last week's in-class coding exercise
  - No worries if you were not able to solve the example exercise during the session
  - It was a rather hard task to get into (no matter how "easy" it was to solve)
- In general: the reference solutions (for the in-class code examples) will now be published sooner

2. Feedback regarding **code** expert

General things regarding code expert

- Re Corrections: I’m on it
- If you **need** the XP: email me

Any questions regarding **code** expert on your part?

3. Learning Objectives

Objectives

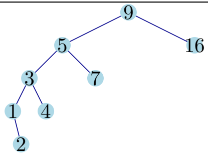
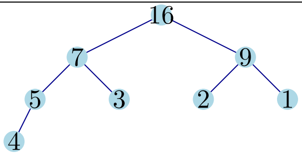
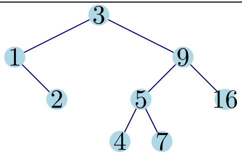
- ☐ Understand *Search Trees* and *Heaps*, and operations on them as well as their drawbacks and benefits
- ☐ Be able to perform operations on *Search Trees* and *Heaps* by hand
- ☐ Understand *Hashing*, its components, and related concepts:
    - ☐ Prehashing
    - ☐ Collision
    - ☐ Simple Uniform Hashing
    - ☐ Uniform Hashing
    - ☐ Open/Closed Addressing & Closed/Open Hashing
    - ☐ Chaining
- ☐ Be able to apply simple *hashing methods* by hand

4. Summary

Getting on the same page

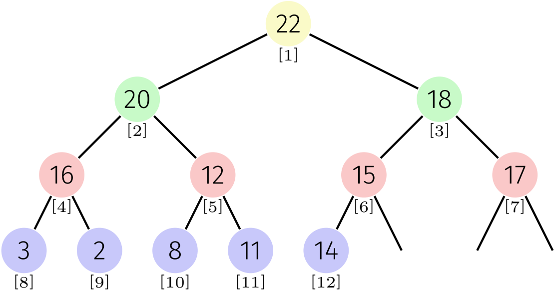
5. Binary Trees and Heaps

Comparison of binary Trees

	Search trees	**Heaps** Min- / Max- Heap	**Balanced trees** AVL, red-black tree
in C++:		`std::make_heap`	`std::map`
			
Insertion	$\Theta(h(T))$	$\Theta(\log n)$	$\Theta(\log n)$
Search	$\Theta(h(T))$	$\Theta(n)$ (!!)	$\Theta(\log n)$
Deletion	$\Theta(h(T))$	Search + $\Theta(\log n)$	$\Theta(\log n)$
Min/Max	$\Theta(h(T))$	$\Theta(1)$/search	$\Theta(\log n)$

Remark: $\Theta(\log n) \leq \Theta(h(T)) \leq \Theta(n)$

Recall: Binary Tree as Array



22	20	18	16	12	15	17	3	2	8	11	14
1	2	3	4	5	6	7	8	9	10	11	12

Repetition: Binary Trees, Inserting a Key

Binary Search Trees

- Search for Key.
- Insert at the reached empty leaf (`null`).

MinHeap

- Insert at the very next free spot (back of the array).
- Restore Heap-Condition: `siftUp` (climb successively).

Exercise: Insert $4, 8, 16, 1, 6, 7$ into empty Search Tree/Min-Heap.

Repetition: Binary Trees, Deleting a Key

Binary Search Trees

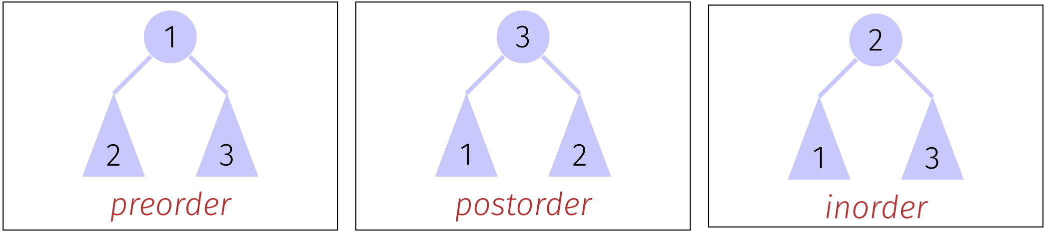
- Replace key k by symmetric successor n.
- Careful: What about right child of n?

MinHeap

- Replace key by last element of the array.
- Restore Heap-Condition: `siftDown` *or* `siftUp`.

Exercise: Delete 4 from Search Tree/Min-Heap.

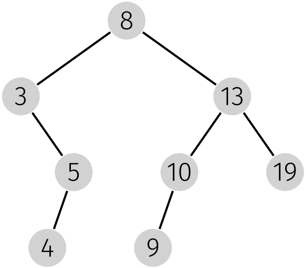
Traversal possibilities



Traversal possibilities

From the Lecture

- *preorder*:
  v, then $T_{\text{left}}(v)$, then $T_{\text{right}}(v)$.
- *postorder*:
  $T_{\text{left}}(v)$, then $T_{\text{right}}(v)$, then v.
- *inorder*:
  $T_{\text{left}}(v)$, then v, then $T_{\text{right}}(v)$.



Quiz

For each of the following traversals, draw a binary search tree that could have produced such a traversal. Is the tree unique, or could different trees have produced this traversal?

inorder	1 2 3 4 5 6 7 8
preorder	4 3 1 2 8 6 5 7
postorder	1 3 2 5 6 8 7 4

Provide for each order a sequence of numbers from $\{1, \ldots, 4\}$ such that it cannot result from a valid binary search tree

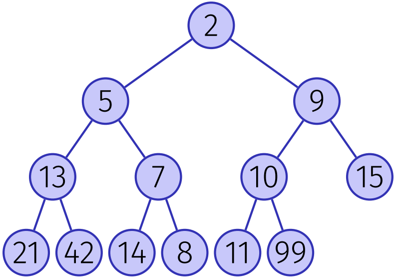
Quiz

True or false:

1. The preorder is the reversed postorder.
2. The first node in the preorder is always the root.
3. The first node in the inorder is never the root.
4. Inserting the nodes in preorder into an empty tree leads to the same tree.
5. Inserting the nodes in postorder into an empty tree leads to the same tree.
6. Inserting the nodes in inorder into an empty tree leads to the same tree.

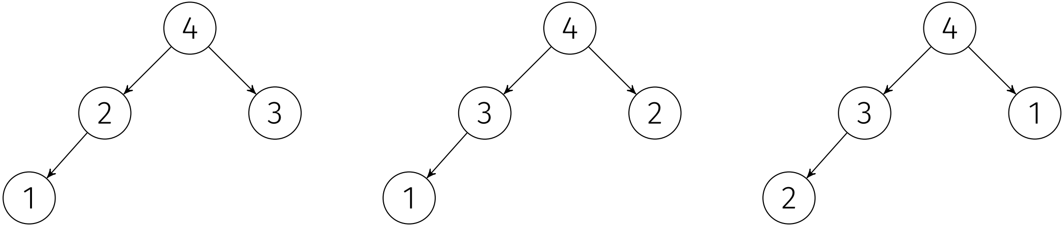
Heap

On the following Min-Heap, perform an extract-min operation, including re-establishing the heap-condition, as shown in class. What does the heap look like after the operation?



Quiz: Number of MaxHeaps on n keys

Let $N(n)$ denote the number of distinct Max-Heaps which can be built from all the keys $1, 2, \ldots, n$. For example we have
$N(1) = 1,\ N(2) = 1,\ N(3) = 2,\ N(4) = 3$ and $N(5) = 8$.
Find the values $N(6)$ and $N(7)$.



6. Hashing

Hashing well-done

Useful Hashing...

- distributes the keys as uniformly as possible in the hash table.
- avoids probing over long areas of used entries (e.g. primary clustering).
- avoids using the same probing sequence for keys with the same hash value (e.g. secondary clustering).

Hashing Examples

Insert the keys $25, 4, 17, 45$ into the hash table, using the function $h(k) = k \bmod 7$ and probing to the right, $h(k) + \mathit{offset}(j, k)$:

- linear probing,
  $\mathit{offset}(j, k) = j$.
- Double Hashing,
  $\mathit{offset}(j, k) = j \cdot (1 + (k \bmod 5))$.

			17	25	4	45

		4	17	25	45	
0	1	2	3	4	5	6

Quiz: Hashing

A hash table of length 10 uses closed hashing with hash function $h(k) = k \bmod 10$, and linear probing (probing goes to the right). After inserting five values into an empty hash table, the table is as shown below.

0	1	2	3	4	5	6	7	8	9
		32	52	33	74	96			

Which of the following *choice(s)* give possible order(s) in which the key values could have been inserted in the table?

(A) 32, 33, 52, 96, 74

(B) 32, 52, 33, 74, 96

(C) 32, 52, 74, 96, 33

(D) 96, 32, 52, 33, 74

7. Binary Tree: Simple Tasks

Hands-on example: Binary Tree

8. Code-Example: Hashtables, Hash-functions and Collisions

Hands-on example: importance of a well designed hashing strategy

9. Past Exam Questions

Past Exam 2022: Task 1d)

In der folgenden Tabelle ist ein Min-Heap in seiner üblichen Form gespeichert. Wie sieht die Tabelle aus, nachdem `ExtractMin` ausgeführt wurde?

The following table comprises a Min-Heap in its canonical form. What does the table look like after `ExtractMin` *has been executed?*

2	5	3	8	7	4	10	13	9	33	11
1	2	3	4	5	6	7	8	9	10	11

1	2	3	4	5	6	7	8	9	10

Past Exam 2022: Task 1d) – Solution

In der folgenden Tabelle ist ein Min-Heap in seiner üblichen Form gespeichert. Wie sieht die Tabelle aus, nachdem `ExtractMin` ausgeführt wurde?

The following table comprises a Min-Heap in its canonical form. What does the table look like after `ExtractMin` *has been executed?*

2	5	3	8	7	4	10	13	9	33	11
1	2	3	4	5	6	7	8	9	10	11

3	5	4	8	7	11	10	13	9	33
1	2	3	4	5	6	7	8	9	10

10. Tips for **code** expert

Finding a Sub-Array

- Given: two integer arrays $A = (a_0, \ldots, a_{n-1})$ and $B = (b_0, \ldots, b_{k-1})$
- Task: Find position of B in A.
- Naive: Loop through A, check whether the following k entries match B.
  - $O(nk)$ comparison operations
- Solution using hashing: Calculate hash $h(B)$ and compare it to $h((a_i, a_{i+1}, \ldots, a_{i+k-1}))$.
- Avoid re-computing $h((a_i, a_{i+1}, \ldots, a_{i+k-1})$ for each $i \implies O(n)$ expected

Sliding Window Hash

- Possible hash function: sum of all elements:
  - Can be updated easily: subtract a_i and add a_{i+k}.
  - However: bad hash function
- Better:

$$H_{c,m}((a_i, \cdots, a_{i+k-1})) = \left(\sum_{j=0}^{k-1} a_{i+j} \cdot c^{k-j-1} \right) \bmod m$$

  - Let c be a prime number: $c = 1021$
  - Let $m = 2^{15} + 3$
  - Since m is just over 2^{15}, intermediate multiplications and additions fit safely in a 32-bit integer, avoiding overflow.

Sliding Window Hash

Make sure that

- the algorithm computes c^k only once,
- all computations are modulo m for all values in order not to get an overflow (recall the rules of modular arithmetic), and
- the values are always positive (e.g., by adding multiples of m).

Computing with Modulo

$$
\begin{aligned}
(a+b) \bmod m &= ((a \bmod m) + (b \bmod m)) \bmod m \\
(a-b) \bmod m &= ((a \bmod m) - (b \bmod m) + m) \bmod m \\
(a \cdot b) \bmod m &= ((a \bmod m) \cdot (b \bmod m)) \bmod m
\end{aligned}
$$

Exercise: Compute

$$12746357 \bmod 11$$

Computing Modulo

Exercise: Compute

$$12746357 \bmod 11$$

11. Outro

General Questions?

See you next time!

Have a nice week!