CSCI 136:
Data Structures
and
Advanced Programming

Lecture 32

Heap implementation

Instructor: Kelly Shaw

Williams

Topics

Heap implementation

Announcements

1. **Final exam**: Saturday, Dec 17, 1:30pm. Room TBD.
2. **Final exam review session**, in class, last day of class, **Friday 12/9**.

Your to-dos

1. **Last quiz**, due **Sat**.
2. Lab 10 (partner lab), **due Tuesday 12/6 by 10pm**.
3. **Review readings** from *Bailey*.
4. **Study** for the final exam.
   a. Pro tip: **review quizzes**.
   b. **Do problems** in study guide/practice exam.
   c. **Don't stress out!** Just be methodical and do your best.
5. **Work on resubmissions** you plan to submit.

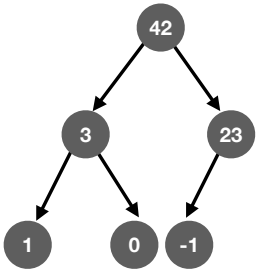
Announcements

Sean Barker '09, Bowdoin College

Friday, Dec 2 @ 2:35pm
Computer Science Colloquium – Wege TCL 123
Smart Meters for Smart Cities: Data Analytics in Energy-Aware Buildings

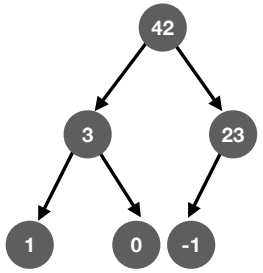
The proliferation of smart energy meters has resulted in many opportunities for next-generation buildings. Energy-aware "smart buildings" may optimize their energy consumption and provide convenience and economic benefits through analysis of their meter data. However, storing and analyzing this data presents computational challenges, especially when conducted at scale. In this talk, I discuss our work on several problems in this space, focusing particularly on efficient compression of smart meter data and the disaggregation of building-wide consumption into individual device consumption. Our work in these areas aims to support the development of sustainable, energy-efficient smart cities and smart grids.

Refresher: binary max heap



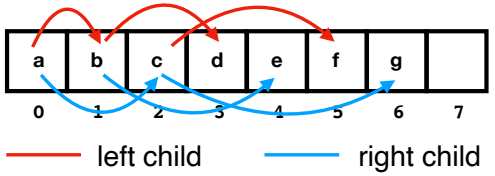
Max heap property: for any given node **n**, if **p** is a parent node of **n**, then the **key** of **p** is ≥ the **key** of **n**.

Insertion



A **binary heap** is usually implemented as an **always-complete binary tree**.

Implementation



A binary heap is often implemented using an implicit binary tree data structure. In other words, heap nodes are actually stored in an array or vector.

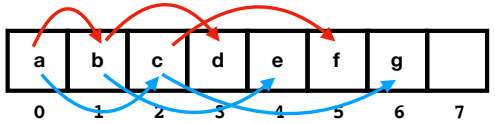
```
leftChild(i) = 2 × i + 1
rightChild(i) = 2 × i + 2
parent(i) = ⌊(i − 1) / 2)⌋
```

Max heap in action

Build a max heap from the following elements:

56 5 57 0 -7 99

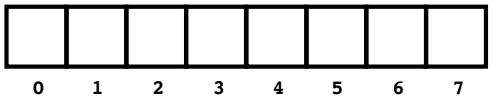
But store the elements in an array (i.e., an implicit binary tree). Process nodes from left to right.



left child right child

```
leftChild(i) = 2 × i + 1
rightChild(i) = 2 × i + 2
parent(i) = ⌊(i − 1) / 2)⌋
```

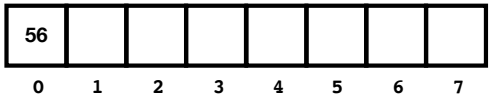
Max heap in action



left child right child

56 5 57 0 -7 99

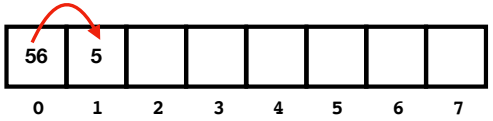
Max heap in action



left child right child

5 57 0 -7 99

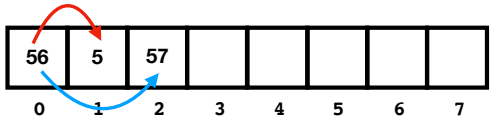
Max heap in action



left child right child

57 0 -7 99

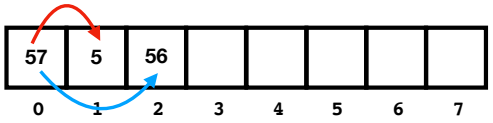
Max heap in action



left child right child

0 -7 99

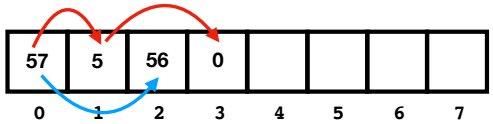
Max heap in action



left child right child

0 -7 99

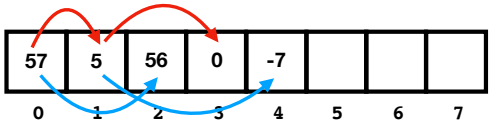
Max heap in action



left child right child

-7 99

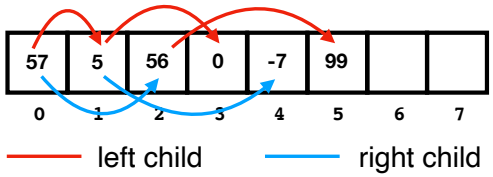
Max heap in action



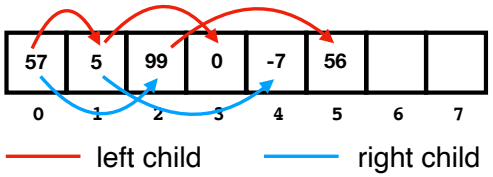
left child right child

99

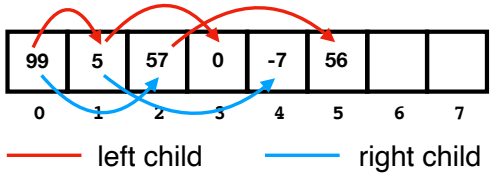
Max heap in action



Max heap in action

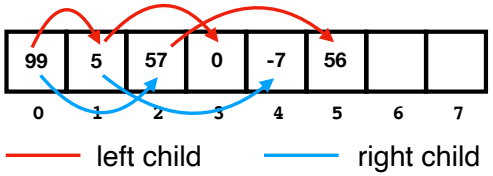


Max heap in action



Done!

Max heap in action



Advantages:

find-max: O(1)
insert: O(log n)
extract: O(log n)

How is a binary heap implemented?
(code)

How about a priority queue?
(code)

Which data structure should I choose?

Activity:
Which data structure should I choose?

For each of the following **scenarios**, **any** of the data structures we've discussed this semester are possible choices.

However, not every data structure is a **good** choice.

Spend some time working with a partner and choose what you think is the **best data structure** for that **scenario**.

Justify your choice by **writing down the reasons** you chose it (hint: asymptotic arguments about space and time are good justifications!)

Activity:

Which data structure should I choose?

a.) You want to count occurrences of each word in a document then print an alphabetical list of word frequencies.

b.) You are writing code that will be used to store thousands of records (e.g. each record contains all the academic information for a specific student) and to retrieve them using a key (e.g. student name). The data rarely changes.

c.) Assume you are given a collection of pairs as input. Each pair contains the names of two people. Taken together, the pairs describe a social network. Over time, you will add pairs when friendships begin and remove them when friendships end. Data changes frequently. At any point in time, you want to be able to find the *k* most "popular" people.

d.) You want to count the occurrences of each letter in a document, then print an alphabetical list of letter frequencies.

Recap & Next Class

Today:

Priority queues

Heaps

Next class:

Dijkstra's algorithm