CSI 3334, Data Structures

Lecture 13: Hashing and Heaps

Date: 2012-10-11
Author(s): Kelleigh Maroney, Will Jones
Lecturer: Fredrik Niemelä

This lecture shows the importance of hashing and heaps. Min/Max, Min/Max/Median, and Max/Min heaps were all discussed.

1 Homework Review

n = of words
m = of table entries
k = of hash functions

Find the probability we have gibberish and think it is a real word.

Answer: $P(k \text{ bits true aka false positive}) = (1 - (1 - \frac{1}{m})^{nk})^k$

Based on the number of hash functions used, you can calculate the size of the table you need to use in order to minimize the probability of a false positive.

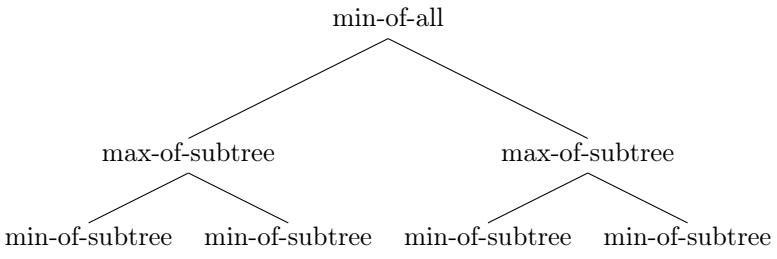
2 Heaps

Heaps are good for priority queues. We looked at Min/Max and Min/Max/Median heaps.

2.1 Min/Max heap

A double ended queue where you can find the minimum and the maximum value in the tree.

The root is the minimum value of the tree. The levels also alternate between mins and maxs of their subtrees.



2.2 Min/Max/Median heap

A triple ended queue where you can find the min, max, and median of the entire tree.

The left subtree is a Min/Max heap. The right subtree is a Max/Min heap. It takes constant time to find and $log(n)$ time to delete because of the use of a "bubble-up."

When something is removed, the median is in the larger subtree so it is neccessary to keep track of the sizes of the two subtrees using a counter. You then need to find the smallest and largest values in that subtree.