CS 106B
Lecture 19: Binary Heaps
Wednesday, November 9, 2016

Programming Abstractions
Fall 2016
Stanford University
Computer Science Department

Lecturer: Chris Gregg

reading:
Programming Abstractions in C++, Chapter ??
1.25 Million people in the U.S. have Type 1 Diabetes
Diabetes is fully treatable, but management is a process that involves closely monitoring blood glucose levels through an under-the-skin sensor, and a trial & error process for delivering insulin through a wearable pump.

The sensor and the pump do not communicate with each other.
Diabetes

Mismanagement can lead to low insulin levels, which can produce nerve and organ damage, seizures, and even death.
However, the FDA has just approved a new device, called a "hybrid closed loop" device where the pump and monitor do communicate. "The device uses advanced algorithms that can learn what the patient’s needs are and can deliver hormones at variable rates."*

This is computer science!

* [http://futurism.com/a-cyborg-society-the-fda-has-approved-the-first-artificial-pancreas/]
But, it isn't all computer science, of course: this device is a technological marvel, and it would not have been possible without the work of scientists, mathematicians, electrical engineers, mechanical engineers, computer scientists, materials scientists, and doctors, among many others. To regulate, market, and distribute the device, businesspeople, lawyers, advertisers, etc., will all be involved.
Working together, humans can be awesome.
Whatever you major in, you have the potential to influence the world for good. Of course, we do have to "learn to walk before we run," and there is a lot to learn in CS before you are developing "advanced algorithms."
Back to Regular Programming: Today's Topics

• Logistics
  • If you used a for loop like this on the exam and got points off, please submit a regrade:
    • `foreach (int x in collection) { ... }`
  • Regrade requests due Friday
• Current assignment: Priority Queue
  • We know you are working hard on the assignments. We have decided to give everyone an extra late day.

• Linked Lists: Deleting a linked list entirely (e.g., in a destructor or clear() function).
• Priority Queues
• Binary Heaps
  • A tree, but not a binary search tree
• The Heap Property
  • Parents have higher priority than children
There was a question on Piazza about destroying a linked list -- this is a key point of understanding, so let's write a linked list clear() function (and call it from the destructor).

The big idea:
- Traverse the list and call delete on each Node
- Don't try to use memory you've already deleted!
• Sometimes, we want to store data in a “prioritized way.”
• Examples in real life:
  • Emergency Room waiting rooms
  • Professor Office Hours (what if a professor walks in? What about the department chair?)
  • Getting on an airplane (First Class and families, then frequent flyers, then by row, etc.)
A “priority queue” stores elements according to their priority, and not in a particular order.
This is fundamentally different from other position-based data structures we have discussed.
There is no external notion of “position.”
Priority Queues

• A priority queue, $P$, has three fundamental operations:

• $\text{enqueue}(k, e)$: insert an element $e$ with key $k$ into $P$.

• $\text{dequeue}()$: removes the element with the highest priority key from $P$.

• $\text{peek}()$: return an element of $P$ with the highest priority key (does not remove from queue).
Priority Queues

• Priority queues also have less fundamental operations:
  • `size()` : returns the number of elements in P.
  • `isEmpty()` : Boolean test if P is empty.
  • `clear()` : empties the queue.
  • `peekPriority()` : Returns the priority of the highest priority element (why might we want this?)
  • `changePriority(string value, int newPriority)` : Changes the priority of a value.
Priority Queues

- Priority queues are simpler than sequences: no need to worry about position (or `insert(index, value)`, `add(value)` to append, `get(index)`, etc.).
- We only need one `enqueue()` and `dequeue()` function
## Priority Queues

<table>
<thead>
<tr>
<th>Operation</th>
<th>Output</th>
<th>Priority Queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>enqueue(5,A)</td>
<td>-</td>
<td>{5,A}</td>
</tr>
<tr>
<td>enqueue(9,C)</td>
<td>-</td>
<td>{5,A},(9,C)</td>
</tr>
<tr>
<td>enqueue(3,B)</td>
<td>-</td>
<td>{5,A},(9,C),(3,B)</td>
</tr>
<tr>
<td>enqueue(7,D)</td>
<td>-</td>
<td>{5,A},(9,C),(3,B),(7,D)</td>
</tr>
<tr>
<td>peek()</td>
<td>B</td>
<td>{5,A},(9,C),(3,B),(7,D)</td>
</tr>
<tr>
<td>peekPriority()</td>
<td>3</td>
<td>{5,A},(9,C),(3,B),(7,D)</td>
</tr>
<tr>
<td>dequeue()</td>
<td>B</td>
<td>{5,A},(9,C),(7,D)</td>
</tr>
<tr>
<td>size()</td>
<td>3</td>
<td>{5,A},(9,C),(7,D)</td>
</tr>
<tr>
<td>peek()</td>
<td>A</td>
<td>{5,A},(9,C),(7,D)</td>
</tr>
<tr>
<td>dequeue()</td>
<td>A</td>
<td>{9,C},(7,D)</td>
</tr>
<tr>
<td>dequeue()</td>
<td>D</td>
<td>{9,C}</td>
</tr>
<tr>
<td>dequeue()</td>
<td>C</td>
<td>{}</td>
</tr>
<tr>
<td>dequeue()</td>
<td>error!</td>
<td>{}</td>
</tr>
<tr>
<td>isEmpty()</td>
<td>TRUE</td>
<td>{}</td>
</tr>
</tbody>
</table>
For HW 5, you will build a priority queue using an unsorted array, a linked list, and ... a "binary heap"

A heap is a *tree-based* structure that satisfies the heap property:
- Parents have a higher priority key than any of their children.
Binary Heaps

- There are two types of heaps:
  - Min Heap
    (root is the smallest element)
  - Max Heap
    (root is the largest element)
binary heaps

- There are no implied orderings between siblings, so both of the trees below are min-heaps:
Circle the min-heap(s):
Binary Heaps

• Circle the min-heap(s):
Heaps are **completely filled**, with the exception of the bottom level. They are, therefore, "complete binary trees":

- complete: all levels filled except the bottom
- binary: two children per node (parent)

- Maximum number of nodes
- Filled from left to right
Binary Heaps

What is the best way to store a heap?

We could use a node-based solution, but…
It turns out that an array works **great** for storing a binary heap!

We will put the root at index 1 instead of index 0 (this makes the math work out just a bit nicer).
Binary Heaps

The array representation makes determining parents and children a matter of simple arithmetic:

- For an element at position \( i \):
  - left child is at \( 2i \)
  - right child is at \( 2i+1 \)
  - parent is at \( \left\lfloor \frac{i}{2} \right\rfloor \)

- \textit{heapSize}: the number of elements in the heap.
Heap Operations

Remember that there are three important priority queue operations:

1. **peek()**: return an element of h with the smallest key.
2. **enqueue(k, e)**: insert an element e with key k into the heap.
3. **dequeue()**: removes the smallest element from h.

We can accomplish this with a heap! We will just look at keys for now -- just know that we will also store a value with the key.
Heap Operations: peek()

peek()

Just return the root!
```
return heap[1]
```

O(1) yay!

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>10</th>
<th>8</th>
<th>12</th>
<th>11</th>
<th>14</th>
<th>13</th>
<th>22</th>
<th>43</th>
</tr>
</thead>
</table>
enqueue\( (k) \)

- How might we go about inserting into a binary heap?

\[
\text{enqueue (9)}
\]

<table>
<thead>
<tr>
<th>5</th>
<th>10</th>
<th>8</th>
<th>12</th>
<th>11</th>
<th>14</th>
<th>13</th>
<th>22</th>
<th>43</th>
</tr>
</thead>
</table>
Heap Operations: enqueue(k)

1. Insert item at element $\text{array}[\text{heap.size()}+1]$ (this probably destroys the heap property)

2. Perform a “bubble up,” or “up-heap” operation:
   a. Compare the added element with its parent — if in correct order, stop
   b. If not, swap and repeat step 2.

See animation at: http://www.cs.usfca.edu/~galles/visualization/Heap.html
Start by inserting the key at the first empty position. This is always at index `heap.size() + 1`.
Heap Operations: enqueue(9)

Start by inserting the key at the first empty position. This is always at index `heap.size()+1`. 
Heap Operations: enqueue(9)

Look at parent of index 10, and compare: do we meet the heap property requirement?

No -- we must swap.
Heap Operations: enqueue(9)
Heap Operations: enqueue(9)
Heap Operations: enqueue(9)

Look at parent of index 5, and compare: do we meet the heap property requirement?

No -- we must swap. This "bubbling up" won't ever be a problem if the heap is "already a heap" (i.e., already meets heap property for all nodes)
Heap Operations: enqueue(9)
Heap Operations: enqueue(9)
Heap Operations: enqueue(9)

No swap necessary between index 2 and its parent. We're done bubbling up!

Complexity? O(log n) - yay!
Average complexity for random inserts: O(1), see: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6312854
Heap Operations: dequeue()

- How might we go about removing the minimum?

\[
dequeue()\
\]

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>9</th>
<th>8</th>
<th>12</th>
<th>10</th>
<th>14</th>
<th>11</th>
<th>22</th>
<th>43</th>
<th>13</th>
</tr>
</thead>
</table>
Heap Operations: dequeue()

1. We are removing the root, and we need to retain a complete tree: replace root with last element.

2. “bubble-down” or “down-heap” the new root:
   a. Compare the root with its children, if in correct order, stop.
   b. If not, swap with smallest child, and repeat step 2.
   c. Be careful to check whether the children exist (if right exists, left must...)

![Heap Diagram]
Heap Operations: dequeue()

```
5
9 8
12 10 14
11
22 43 13
```

```
```
Heap Operations: dequeue()

Remove root (will return at the end)

5

\[
\begin{array}{ccccccccccc}
5 & 9 & 8 & 12 & 10 & 14 & 11 & 22 & 43 & 13 \\
\end{array}
\]
Heap Operations: dequeue()

Move last element (at `heap[heap.size()]`) to the root (this may be unintuitive!) to begin bubble-down

Don't forget to decrease heap size!
Heap Operations: dequeue()

Compare children of root with root: swap root with the smaller one (why?)
Heap Operations: dequeue()

Keep swapping new element if necessary. In this case: compare 13 to 11 and 14, and swap with smallest (11).
13 has now bubbled down until it has no more children, so we are done!

Complexity? O(log n) - yay!
Heaps in Real Life

- Heapsort (see extra slides)
- Google Maps -- finding the shortest path between places
- All priority queue situations
- Kernel process scheduling
- Event simulation
- Huffman coding
What is the best method for building a heap from scratch (buildHeap())

14, 9, 13, 43, 10, 8, 11, 22, 12

We could insert each in turn.
An insertion takes $O(\log n)$, and we have to insert $n$ elements

Big O? $O(n \log n)$
Heap Operations: building a heap from scratch

There is a better way: **heapify()**
1. Insert all elements into a binary tree in original order (O(n))

2. Starting from the lowest completely filled level at the first node with children (e.g., at position n/2), down-heap each element (also O(n) to heapify the whole tree).

```java
for (int i=heapSize/2;i>0;i--){
    downHeap(i);
}
```
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

first node with children!

loop down:
i = heapSize/2
heapSize == 9,
i == 4
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

i==4
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

i==3
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

no swap necessary

14, 9, 13, 43, 10, 8, 11, 22, 12

i==2
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

i == 1
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

must keep down-heaping
Heap Operations: building a heap from scratch

14, 9, 13, 43, 10, 8, 11, 22, 12

Done!
We now have a proper min-heap.
Asymptotic complexity — not trivial to determine, but turns out to be O(n).
Heap Operations: heaping: empirical

BuildHeap

Empirical Results

(Java)
Heap Operations: heaping: empirical

BuildHeap Empirical Results (C++)

![Graph showing BuildHeap (C++) time vs. number of elements]
References and Advanced Reading

• References:
  • Priority Queues, Wikipedia: http://en.wikipedia.org/wiki/Priority_queue
  • YouTube on Priority Queues: https://www.youtube.com/watch?v=gJc-J7K_P_w
  • http://en.wikipedia.org/wiki/Binary_heap (excellent)
  • http://www.cs.usfca.edu/~galles/visualization/Heap.html (excellent visualization)
  • Another explanation online: http://www.cs.cmu.edu/~adamchik/15-121/lectures/Binary%20Heaps/heaps.html (excellent)

• Advanced Reading:
  • YouTube video with more detail and math: https://www.youtube.com/watch?v=B7hVxCmfPtM (excellent, mostly max heaps)
Extra Slides
Extras: HeapSort

• We can perform a full heap sort in place, in $O(n \log n)$ time.
• First, heapify an array (i.e., call build-heap on an unsorted array)
• Second, iterate over the array and perform dequeue(), but instead of
  returning the minimum elements, swap them with the last element (and
  also decrease heapSize)
• When the iteration is complete, the array will be sorted from low to high
  priority.
Extras: HeapSort — Heapify first

Unheaped:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>5</td>
<td>11</td>
<td>8</td>
<td>12</td>
<td>10</td>
<td>14</td>
<td>2</td>
<td>22</td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heaped:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>11</td>
<td>14</td>
<td>9</td>
<td>22</td>
<td>43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>43</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>11</td>
<td>14</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td>3</td>
<td></td>
<td>4</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

heapSize

---

1. **HeapSort**: A sorting algorithm that takes advantage of a binary heap data structure. It maintains a max heap, where the root node is the largest element. The algorithm repeatedly removes the root (the largest element) and replaces it with the last element in the heap, then reheapifies the root to maintain the max heap property.
2. **Iterate and dequeue()**: In the context of HeapSort, after removing the root (largest element), the next largest element is removed and swapped with the last element in the heap. This process is repeated until all elements are sorted.
3. **Down-heaping**: After removing the root, the heap size decreases. The remaining elements are reorganized to maintain the max heap property.
Extras: HeapSort — Iterate and call dequeue(), swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call dequeue(), swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with
the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call dequeue(), swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call **dequeue()**, swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call \texttt{dequeue()}\texttt{()}, swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[0]</td>
<td>22</td>
<td>12</td>
<td>43</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>9</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

Heap Sort

```
heapSize
```

```

<table>
<thead>
<tr>
<th>12</th>
<th>43</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>
```

```
Extras: HeapSort — Iterate and call dequeue(), swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call \texttt{dequeue()}\index{dequeue!function} , swapping the root with the last element, then down-heaping.
Extras: HeapSort — Iterate and call `dequeue()` , swapping the root with the last element, then down-heaping.

Done! (reverse-ordered)

Complexity: $O(n \log n)$
HeapSort Empirical Results (Java)
HeapSort Empirical Results (C++)

The graph shows the time (ms) on the y-axis and the number of elements on the x-axis. The lines represent different time complexities:

- Blue line: $n \log n$
- Green line: $n$
- Red line: Empirical

The empirical results are compared to the theoretical complexities $n \log n$ and $n$. The graph illustrates how the time increases with the number of elements, with the empirical results closely following the $n \log n$ line, indicating an efficient performance.
Consider a full binary heap data structure with \( n \) nodes.

Nodes at this level: 1, work done: \( c \times (1) \times \log n \)

Nodes at this level: \( n/8 \), work done: \( c \times n/8 \times 2 \)

Nodes at this level: \( n/4 \), work done: \( c \times n/4 \times 1 \) (possible swaps to bottom level)

Work at this level: none

Extras: Why is \( \text{buildheap()} \) \( O(n) \)?
Consider a full binary heap data structure with $n$ nodes.

Total work done:

$$c \frac{n}{4} \cdot 1 + c \frac{n}{8} \cdot 2 + c \frac{n}{16} \cdot 3 + \cdots + c(1) \cdot \log(n)$$

Extras: Why is \texttt{buildheap()} $O(n)$?
Consider a full binary heap data structure with n nodes.

Total work done:
\[ \sum k \cdot \frac{n}{c^k} \]

Substitution: \[ \frac{n}{4} = 2^k \]

Must do some math for \( \lg(n) \):
\[
\begin{align*}
n &= 4 \cdot 2^k = 2^2 \cdot 2^k = 2^{k+2} \\
\lg(n) &= \lg(2^{k+2}) = k + 2
\end{align*}
\]
Consider a full binary heap data structure with \( n \) nodes.

With substitution, and pulling out \( c \cdot 2^k \):

\[
c \cdot 2^k \left( \frac{1}{2^0} + \frac{2}{2^1} + \frac{3}{2^2} + \cdots + \frac{k + 2}{2^k} \right)
\]

Simplify a bit more:

\[
\frac{k + 2}{2^k} = \frac{k + 1 + 1}{2^k} = \frac{k + 1}{2^k} + \frac{1}{2^k}
\]

Extras: Why is buildheap() \( O(n) \)?
Consider a full binary heap data structure with \( n \) nodes.

With substitution, and pulling out \( c \cdot 2^k \):

\[
c \cdot 2^k \left( \frac{1}{2^0} + \frac{2}{2^1} + \frac{3}{2^2} + \cdots + \frac{k + 2}{2^k} \right)
\]

\[
c \cdot 2^k \left( \sum_{i=0}^{k} \frac{i + 1}{2^i} + \frac{1}{2^k} \right)
\]

\[
\sum_{i=0}^{k} \frac{i + 1}{2^i} = 4
\]
Consider a full binary heap data structure with $n$ nodes. With substitution, and pulling out $c \times 2^k$:

$$c \times 2^k \left(4 + \frac{1}{2^k}\right)$$

Substitution:

$$\frac{n}{4} = 2^k$$

$$c \times n + c \quad \text{Linear amount of work!}$$