CSE5311 Design and Analysis of Algorithms

• This Class
  – Heaps and Heapsort?
  – QuickSort
  – Mergesort
  – Other Sorting Algorithms

• At the end of the class
  – Binary trees
  – Priority queues and heaps
  – Quicksort
    – Worstcase
    – Bestcase
  – Mergesort
  – Recurrences for Quicksort and Mergesort

Further Reading
Reference books on Algorithms

Course Syllabus

• Review of Asymptotic Analysis and Growth of Functions, Recurrences

• Sorting Algorithms
  • Graphs and Graph Algorithms.
  • Greedy Algorithms:
    – Minimum spanning tree, Union-Find algorithms, Kruskal’s Algorithm,
    – Clustering,
    – Huffman Codes, and
    – Multiphase greedy algorithms.
  • Dynamic Programming:
    – Shortest paths, negative cycles, matrix chain multiplications, sequence alignment, RNA secondary structure, application examples.

• Network Flow:

• NP and Computational tractability:
  – Polynomial time reductions; The Satisfiability problem; NP-Complete problems; and Extending limits of tractability.

• Approximation Algorithms, Local Search and Randomized Algorithms
Priority Queues

What is a priority queue?
A priority queue is an abstract data type which consists of a set of elements. Each element of the set has an associated priority or key.
Priority is the value of the element or value of some component of an element.

Example:
S : {(Brown, 20), (Gray, 22), (Green, 21)} priority based on name
{(Brown, 20), (Green, 21), (Gray, 22)} priority based on age

Each element could be a record and the priority could be based on one of the fields of the record.
Example

A Student's record:

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Name</th>
<th>Age</th>
<th>Sex</th>
<th>Student No.</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>John Brown</td>
<td>21</td>
<td>M</td>
<td>94XYZ23</td>
<td>75</td>
</tr>
</tbody>
</table>

Priority can be based on name, age, student number, or marks

Operations performed on priority queues,
- inserting an element into the set
- finding and deleting from the set an element of highest priority

Priority Queues

Priority queues are implemented on partially ordered trees (POTs)
- POTs are labeled binary trees
- the labels of the nodes are elements with a priority
- the element stored at a node has at least as large a priority as the elements stored at the children of that node
- the element with the highest priority is at the root of the tree
HEAPS

The heap is a data structure for implementing POT's. Each node of the heap tree corresponds to an element of the array that stores the value in the node. The tree is filled on all levels except possibly the lowest, which are filled from left to right up to a point.

An array A that represents a heap is an object with two attributes

- length[A], the number of elements in the array and
- heap-size[A], the number of elements in the heap stored within the array A

heap_size[A] ≤ length[A]
HEAPS (Contd)


<table>
<thead>
<tr>
<th>Position</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>Value</td>
<td>24</td>
<td>21</td>
<td>19</td>
<td>13</td>
<td>14</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

Given node with index $i$,

PARENT($i$) is the index of parent of $i$; PARENT($i$) = $\lceil i/2 \rceil$

LEFT_CHILD($i$) is the index of left child of $i$;
LEFT_CHILD($i$) = $2i$;

RIGHT_CHILD($i$) is the index of right child of $i$; and
RIGHT_CHILD($i$) = $2i + 1$

Heap Property

THE HEAP PROPERTY
$A[\text{PARENT}(i)] \geq A[i]$

The heap is based on a binary tree
The height of the heap (as a binary tree) is the number of edges on the longest simple downward path from the root to a leaf.

The height of a heap with $n$ nodes is $O(\log n)$.

All basic operations on heaps run in $O(\log n)$ time.
\[ n = 2^0 + 2^1 + 2^2 + 2^3 + \ldots + 2^h = 2^{h+1} - 1 \]

Number of nodes at different levels in a Binary Tree

Heap Algorithms

- HEAPIFY
- BUILD_HEAP
- HEAPSORT
- HEAP_EXTRACT_MAX
- HEAP_INSERT
The HEAPIFY algorithm checks the heap elements for violation of the heap property and restores heap property.

Procedure HEAPIFY (A, i)
Input: An array A and index i to the array. i = 1 if we want to heapify the whole tree. Subtrees rooted at LEFT_CHILD(i) and RIGHT_CHILD(i) are heaps
Output: The elements of array A forming subtree rooted at i satisfy the heap property.

1. \( l \leftarrow \text{LEFT_CHILD}(i); \)
2. \( r \leftarrow \text{RIGHT_CHILD}(i); \)
3. if \( l \leq \text{heap_size}[A] \) and \( A[l] > A[i] \)
   then largest \( \leftarrow l; \)
   else largest \( \leftarrow i; \)
4. if \( r \leq \text{heap_size}[A] \) and \( A[r] > A[\text{largest}] \)
   then largest \( \leftarrow r; \)
5. if largest \( \neq i \)
6. then exchange \( A[i] \leftrightarrow A[\text{largest}] \)
7. HEAPIFY (A, largest)
Running time of HEAPIFY

Total running time = steps 1 … 9 + recursive call
\[ T(n) = \Theta(1) + T(n/2) \]
Solving the recurrence, we get \( T(n) = O(\log n) \)

BUILD_HEAP

Procedure BUILD_HEAP (A)
\begin{itemize}
  \item [Input] An array A of size \( n = \text{length}[A] \), \( \text{heap}_{\text{size}}[A] \)
  \item [Output] A heap of size n
\end{itemize}
\begin{enumerate}
  \item \( \text{heap}_{\text{size}}[A] \leftarrow \text{length}[A] \)
  \item for \( i \leftarrow \left\lfloor \text{length}[A]/2 \right\rfloor \) downto 1
  \item \( \text{HEAPIFY}(A, i) \)
\end{enumerate}
Running time of Build_heap

1. Each call to HEAPIFY takes $O(\log n)$ time
2. There are $O(n)$ such calls
3. Therefore the running time is at most $O(n \log n)$
   However the complexity of BUILD_HEAP is $O(n)$

Proof:
In an n element heap there are at most $\left\lceil \frac{n}{2^h+1} \right\rceil$ nodes of height $h$
The time required to heapify a subtree whose root is at a height $h$ is $O(h)$
   (this was proved in the analysis for HEAPIFY)
So the total time taken for BUILD_HEAP is given by,

$$\leq \sum_{h=0}^{\log n} \left\lceil \frac{n}{2^{h+1}} \right\rceil \cdot h$$

$$\leq \frac{n}{2} \cdot \sum_{h=0}^{\log n} \frac{h}{2^h}$$

We know that

$$\sum_{h=0}^{\infty} \frac{h}{2^h} = 2 = O(n)$$

Thus the running time of BUILD_HEAP is given by, $O(n)$
The HEAPSORT Algorithm

Procedure HEAPSORT(A)
Input : Array A[1…n], n = length[A]
Output : Sorted array A[1…n]
1. BUILD_HEAP[A]
2. for i ← length[A] down to 2
4. heap_size[A] ← heap_size[A]-1;
5. HEAPIFY(A,1)

Example : To be given in the lecture

HEAPSORT

Running Time:
Step 1 BUILD_HEAP takes O(n) time,
Steps 2 to 5 : there are (n-1) calls to HEAPIFY
which takes O(log n) time
Therefore running time takes  O (n log n)
HEAP_EXTRACT_MAX

Procedure HEAP_EXTRACT_MAX(A[1…n])
Input : heap(A)
Output : The maximum element or root, heap (A[1…n-1])
1. if heap_size[A] ≥ 1
2. max ← A[1];
4. heap_size[A] ← heap_size[A]-1;
5. HEAPIFY(A,1)
6. return max

Running Time : O (log n) time

HEAP_INSERT

Procedure HEAP_INSERT(A, key)
Input : heap(A[1…n]), key - the new element
Output : heap(A[1…n+1]) with k in the heap
1. heap_size[A] ← heap_size[A]+1;
2. i ← heap_size[A];
3. while i > 1 and A[PARENT(i)] < key
5. i ← PARENT(i);
6. A[i] ← key

Running Time : O (log n) time
Questions:

- What is a heap?
- What are the running times for heap insertion and deletion operations?
- Did you understand HEAPIFY AND and HEAPSORT algorithms?
- Can you write a heapsort algorithm for arranging an array of numbers in descending order?

Quicksort

- Quicksort algorithm
- Quicksort performance
- Quicksort analysis
## Quicksort

- The worst case running time of Quicksort algorithm is \( O(n^2) \)
- However, its expected running time is \( O(n \log n) \)
- Three-step divide-and-conquer process for sorting a subarray \( A[l..r] \)

**Divide**: partition the array \( A[l..r] \) into two nonempty subarrays \( A[l..q] \) and \( A[q+1..r] \) such that each element of \( A[l..q] \) is less than or equal to each element of \( A[q+1..r] \)

**Conquer**: sort the two subarrays \( A[l..q] \) and \( A[q+1..r] \) by recursive calls to Quicksort

**Combine**: the subarrays are already sorted in place. No work is needed to combine them

### Example

<table>
<thead>
<tr>
<th>13</th>
<th>02</th>
<th>18</th>
<th>26</th>
<th>76</th>
<th>87</th>
<th>98</th>
<th>11</th>
<th>93</th>
<th>77</th>
<th>65</th>
<th>43</th>
<th>38</th>
<th>09</th>
<th>65</th>
<th>06</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>02</td>
<td>06</td>
<td>26</td>
<td>76</td>
<td>87</td>
<td>98</td>
<td>11</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>09</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>13</td>
<td>02</td>
<td>06</td>
<td>09</td>
<td>76</td>
<td>87</td>
<td>98</td>
<td>11</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>02</td>
<td>06</td>
<td>09</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>09</td>
<td>02</td>
<td>06</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>09</td>
<td>02</td>
<td>06</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>06</td>
<td>02</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>06</td>
<td>02</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>06</td>
<td>02</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>09</td>
<td>02</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>06</td>
<td>02</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>98</td>
<td>76</td>
<td>93</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>87</td>
<td>18</td>
<td>76</td>
<td>65</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>26</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>26</td>
<td>18</td>
<td>76</td>
<td>65</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>26</td>
<td>18</td>
<td>76</td>
<td>65</td>
<td>77</td>
<td>65</td>
<td>43</td>
<td>38</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>76</td>
<td>65</td>
<td>38</td>
<td>65</td>
<td>43</td>
<td>77</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>43</td>
<td>65</td>
<td>38</td>
<td>65</td>
<td>76</td>
<td>77</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>43</td>
<td>38</td>
<td>65</td>
<td>65</td>
<td>76</td>
<td>77</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>38</td>
<td>43</td>
<td>65</td>
<td>65</td>
<td>76</td>
<td>77</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>02</td>
<td>06</td>
<td>09</td>
<td>11</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>38</td>
<td>43</td>
<td>65</td>
<td>65</td>
<td>76</td>
<td>77</td>
<td>87</td>
<td>93</td>
<td>98</td>
</tr>
</tbody>
</table>
Procedure Quicksort(A,\(l,r\))

**Input**: Unsorted Array (A,\(l,r\))

**Output**: Sorted subarray A(0..r)

To sort the entire array A, \(l = 1\) and \(r = \text{length}[A]\)

1. if \(l < r\)
   2. \(q \leftarrow \text{PARTITION}(A, l, r)\)
   3. QUICKSORT(A, \(l,q-1\))
   4. QUICKSORT(A, \(q+1,r\))
Running Time of Quicksort

\[ T(n) = n - 1 + T(i - 1) + T(n - i) \]

It takes \( n - 1 \) comparisons for the partition
Then we sort smaller sequences of size \( i - 1 \) and \( n - i \)
Each element has the same probability of being selected as
the pivot,
The average running time is given by,

\[ T(n) = n - 1 + \frac{1}{n} \sum_{i=1}^{n} T(i - 1) + \frac{1}{n} \sum_{i=1}^{n} T(n - i) \]

\[ T(n) = n - 1 + \frac{2}{n} \sum_{i=1}^{n} T(i) = O(n \log n) \]

Detailed Analysis – ? Discussion

Mergesort
Mergesort

Like Quicksort, Mergesort algorithm also is based on the divide-and-conquer principle.

Divide: This step computes the middle of the array, takes constant time, $\Theta(1)$

Conquer: Two subproblems, each of size $n/2$ are recursively solved. Each subproblem contributes $2T(n/2)$ to the running time.

Combine: Two sorted sequences are merged, this takes $\Theta(n)$ time

Example

13 02 18 26 76 87 98 11 93 77 65 43 38 09 65 06
13 02 18 26 76 87 98 11 93 77 65 43 38 09 65 06
13 02 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 13 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 13 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 13 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 13 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 13 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 13 18 26 76 87 98 11 93 77 65 43 38 09 65 06
02 11 13 18 26 76 87 98 93 77 65 43 38 09 65 06
02 11 13 18 26 76 87 98 93 77 65 43 38 09 65 06
02 06 09 11 13 18 26 38 43 65 65 77 77 87 93 98
Procedure MERGESORT(A, l, r)
Input: A an array in the range 1 to n.
Output: Sorted array A.
if l < r
then q ← \lfloor (l + r)/2 \rfloor;
    MERGESORT(A, l, q)
    MERGESORT(A, q+1, r)
MERGE (A, l, q, r)

Procedure MERGE(A, l, q, r)
Inputs: two sorted subarrays A(l, q) and A(q+1, r)
Output: Merged and sorted array A(l, r)
i ← l;
j ← q+1;
k ← 0;
while (i ≤ q) and (j ≤ r) do
    k ← k+1;
    if A[i] ≤ A[j] then
        TEMP[k] ← A[i];
        i ← i +1;
    else
        TEMP[k] ← A[j];
        j ← j +1;
if j > r then
    for t ← 0 to q – i do
        A[r-t] ← A[q-t];
for t ← 0 to k-1 do
    A[l +t] ← TEMP[t];
Runtime Complexity of Mergesort

\[ T(n) = 2 T(n/2) + \Theta(n) \]
\[ T(n/2) = 2 T(n/4) + n/2 \]
\[ T(n/4) = 2 T(n/8) + n/4 \]
\[ T(n) = 2(2 T(n/4) + n/2) + n = 4 T(n/4) + 2 n \]
\[ T(n) = 2^3 T(n/2^3) + 3n \]
\[ \ldots \]
\[ T(n) = 2^k T(n/2^k) + k n \text{ if } n = 2^k \text{ then } , k = \log_2 n \]

Therefore, \( T(n) = 2^k T(1) + n \log n \)
\[ = n \Theta(1) + n \log n \]
\[ T(n) = O(n \log n) \]
Questions

- What is a pivot element?
- Do you understand divide-and-conquer?
- What is running time if pivot element is at the center of the array?
- How is Mergesort different from Quicksort?
- Trace the algorithm with the help of an example?
- What is the use of TEMP array?
- Do you know why we execute steps 13 and 14 in the MERGE algorithm?
- What happens if i > q after the while loop (4-11)?

Home Work

- Find applications for selection sort, heapsort, quicksort, and mergesort sorting algorithms?
- Which of these problems are suitable for different types of sorting operations?
- Identify applications for which each sorting algorithm works best
Homework - Reference Books

• Insertion Sort
• Counting Sort
• Find the Maximum
• Find the Minimum