Binary Trees &
Binary Search Trees

(data structures for the dictionary ADT)
Outline

- Binary tree terminology
- Tree traversals: preorder, inorder and postorder
- Dictionary and binary search tree
- Binary search tree operations
  - Search
  - min and max
  - Successor
  - Insertion
  - Deletion
- Tree balancing issue
Binary Tree Terminology

- Go to the supplementary notes
The degree of a node is the number of children it has. The degree of a tree is the maximum of its element degree.

- In a binary tree, the tree degree is two.

Each node has two links:
- one to the left child of the node
- one to the right child of the node
- if no child node exists for a node, the link is set to NULL.
Binary Trees as Recursive Data Structures

- A binary tree is either empty ... or
- Consists of a node called the root
- Root points to two disjoint binary (sub)trees left and right (sub)tree

COMP2012H (BST)
Tree Traversal is Also Recursive (Preorder example)

If the binary tree is empty then
   do nothing

Else
   N: Visit the root, process data
   L: Traverse the left subtree
   R: Traverse the right subtree

Anchor

Inductive/Recursive step
3 Types of Tree Traversal

- If the pointer to the node is not NULL:
  - **Preorder**: Node, Left subtree, Right subtree
  - **Inorder**: Left subtree, Node, Right subtree
  - **Postorder**: Left subtree, Right subtree, Node

```cpp
template<class T>
void BinaryTree<T>::PreOrder(
    void(*Visit)(BinaryTreeNode<T> *u),
    BinaryTreeNode<T> *t)
{// Preorder traversal.
    if (t) {
        Visit(t);
        PreOrder(Visit, t->LeftChild);
        PreOrder(Visit, t->RightChild);
    }
}
```

```cpp
template <class T>
void BinaryTree<T>::InOrder(
    void(*Visit)(BinaryTreeNode<T> *u),
    BinaryTreeNode<T> *t)
{// Inorder traversal.
    if (t) {
        InOrder(Visit, t->LeftChild);
        Visit(t);
        InOrder(Visit, t->RightChild);
    }
}
```

```cpp
template <class T>
void BinaryTree<T>::PostOrder(
    void(*Visit)(BinaryTreeNode<T> *u),
    BinaryTreeNode<T> *t)
{// Postorder traversal.
    if (t) {
        PostOrder(Visit, t->LeftChild);
        PostOrder(Visit, t->RightChild);
        Visit(t);
    }
}
```
Traversals Order

- Given expression
  \[ A - B \times C + D \]
- Child node: operand
- Parent node: corresponding operator
- Inorder traversal: infix expression
  \[ A - B \times C + D \]
- Preorder traversal: prefix expression
  \[ + - A \times B \times C \times D \]
- Postorder traversal: postfix or RPN expression
  \[ A \times B \times C - D + \]
Preorder, Inorder and Postorder Traversals

Figure 8.5 Expression trees

Preorder  \(+ab/cd\  \;+++abcd\  \;/+-a+xy*+b*ca\)
Inorder   \(a*b+c/d\ a+b+c+d\  -a+x+y+/+b*c*a\)
Postorder \(ab*cd/+\  \;ab+c+d\  \;a-xy++b+ca**/\)

Figure 8.11 Elements of a binary tree listed in pre-, in-, and postorder
A Faster Way for Tree Traversal

- You may eye-ball the solution without using recursion.
- First emanating from each node a “hook.” Trace from left to right an outer envelop of the tree starting from the root. Whenever you touch a hook, you print out the node.
- **Preorder:**
  - put the hook to the left of the node
- **Inorder:**
  - put the hook vertically down at the node
- **Postorder:**
  - put the hook to the right of the node
Another Example (This is a Search Tree)

- Inorder (Left, Visit, Right): 2, 3, 4, 6, 7, 9, 13, 15, 17, 18, 20
- Preorder (Visit, Left, Right): 15, 6, 3, 2, 4, 7, 13, 9, 18, 17, 20
- Postorder (Left, Right, Visit): 2, 4, 3, 9, 13, 7, 6, 17, 20, 18, 15
template <class T>
void Infix(BinaryTreeNode<T> *t)
{// Output infix form of expression.
    if (t) {
        cout << '(';
        Infix(t->LeftChild);   // left operand
        cout << t->data;       // operator
        Infix(t->RightChild);  // right operand
        cout << ')';
    }
}

+   returns ((a)+(b))
a  b
Infix to Prefix (Pre-order Expressions)

- Infix = In-order expression

1. Infix to postfix
2. postfix to build an expression tree
   1. Push operands into a stack
   2. If an operator is encountered, create a binary node with the operator as the root, push once as right child, push the 2\textsuperscript{nd} time as left child, and push the complete tree into the stack
3. With the expression tree, traverse in preorder manner
   - Parent-left-right

COMP2012H (BST)
Binary Search Tree
Linear Search on a Sorted Sequence

- Collection of ordered data items to be searched is organized in a list
  \[ x_1, x_2, \ldots, x_n \]

- Assume `==` and `<` operators defined for the type

- Linear search begins with item 1
  - continue through the list until target found
  - or reach end of list
template <typename t>
void LinearSearch (const vector<t> &v, const t &item,
                  boolean &found, int &loc)
{
    found = false;  loc = 0;
    for (; ; )
    {
        if (found || loc == v.size())
            return;
        if (item == v[loc])
            found = true;
        else
            loc++;
    }
}
Binary Search: Vector Based

```cpp
template <typename t>
void LinearSearch (const vector<t> &v, const t &item, 
                   boolean &found, int &loc) 
{
    found = false;
    int first = 0;
    last = v.size() - 1;
    for ( ; ; )
    {
        if (found || first > last) return;
        loc = (first + last) / 2;
        if (item < v[loc])
            last = loc - 1;
        else if (item > v[loc])
            first = loc + 1;
        else
            /* item == v[loc] */
            found = true;
    }
}
```

May be replaced by recursive codes with additional function parameters first and last.
Binary Search

- Outperforms a linear search (infinitely faster asymptotically)
- Disadvantage:
  - Requires a sequential storage
  - Not appropriate for linked lists (Why?)
- It is possible to use a linked structure which can be searched in a binary-like manner
  - Binary tree
Dictionary

- A dictionary is a collection of elements
- Each element has a field called key
- No two elements have the same key value

AbstractDataType Dictionary {
    instances
        collection of elements with distinct keys
    Operations
        Create (): create an empty dictionary
        Search (k,x): return element with key k in x;
            return false if the operation
            fails, true if it succeeds
        Insert (x): insert x into the dictionary
        Delete (k,x): delete element with key k and
            return it in x
}

COMP2012H (BST)
Binary Search Tree (BST)

- Collection of data elements in a binary tree structure
- Stores keys in the nodes of the binary tree in a way so that searching, insertion and deletion can be done efficiently
- Every element has a key (or value) and no two elements have the same key (all keys are distinct)
- The keys (if any) in the left subtree of the root are smaller than the key in the root
- The keys (if any) in the right subtree of the root are larger than the key in the root
- The left and right subtrees of the root are also binary search trees
Binary Search Tree

for any node $y$ in this subtree
\[ \text{key}(y) < \text{key}(x) \]

for any node $z$ in this subtree
\[ \text{key}(z) > \text{key}(x) \]
Examples of BST

- For each node \( x \),
  values in left subtree \( \leq \) value in \( x \) \( \leq \) value in right subtree
- a) is NOT a search tree, b) and c) are search trees
Binary Search Tree Property

- Two binary search trees representing the same set
Sorting: Inorder Traversal for a Search Tree

- Print out the keys in sorted order

- A simple strategy is to
  1. print out all keys in left subtree in sorted order;
  2. print 15;
  3. print out all keys in right subtree in sorted order;
Indexed Binary Search Tree

- Derived from binary search tree by adding another field $LeftSize$ to each tree node
- $LeftSize$ gives the number of elements in the node’s left subtree plus one
- An example (the number inside a node is the element key, while that outside is the value of $LeftSize$)
- It is the rank of the node for the search tree rooted at that node (rank is the position in the sorted order)
- Can be used to figure out the rank of the node in the tree
Tree Search

- If we are searching for 15, then we are done
- If we are searching for a key < 15, then we should search for it in the left subtree
- If we are searching for a key > 15, then we should search for it in the right subtree
An Example

Search for 9:
1. compare 9:15 (the root), go to left subtree;
2. compare 9:6, go to right subtree;
3. compare 9:7, go to right subtree;
4. compare 9:13, go to left subtree;
5. compare 9:9, found it!
Find Min and Max

Minimum element is always the left-most node.

Maximum element is always the right-most node.

Algorithm \textit{Minimum}(x)
\begin{itemize}
  \item \textbf{Input:} \(x\) is the root.
  \item \textbf{Output:} the node containing the minimum key.
  \item 1. \textbf{while} \(\text{left}(x) \neq \text{NULL}\)
  \item 2. \textbf{do} \(x := \text{left}(x)\);
  \item 3. \textbf{return} \(x\);
\end{itemize}

Algorithm \textit{Maximum}(x)
\begin{itemize}
  \item \textbf{Input:} \(x\) is the root.
  \item \textbf{Output:} the node containing the maximum key.
  \item 1. \textbf{while} \(\text{right}(x) \neq \text{NULL}\)
  \item 2. \textbf{do} \(x := \text{right}(x)\);
  \item 3. \textbf{return} \(x\);
\end{itemize}

Time Complexity
Worse case?
Height of tree, which can be the total number of nodes if tree is not balanced!
Successor

The successor of a node $x$ is defined as:

- The node $y$, whose key($y$) is the successor of key($x$) in sorted order
  of this tree. (2,3,4,6,7,9,13,15,17,18,20)

Some examples:

Which node is the successor of 2?
Which node is the successor of 9?
Which node is the successor of 13?
Which node is the successor of 20? Null

Successor of 13
Successor of 6
Successor of 9
Finding Successor:
Three Scenarios to Determine Successor

Successor(x)

- **Scenario I**: x has right descendants
  \[ \Rightarrow \text{minimum}(\text{right}(x)) \]

- **Scenario II**: x is the *left* child of some node
  \[ \Rightarrow \text{parent}(x) \]

- **Scenario III**: x is the *right* child of some node
Scenario I: Node $x$ Has a Right Subtree

By definition of BST, all items greater than $x$ are in this right sub-tree.

Successor is the minimum( right( $x$ ) )

Scenario I
Scenario II: Node $x$ Has No Right Subtree and $x$ is the Left Child of Parent (x)

Successor is $\text{parent}(x)$

Why? The successor is the node whose key would appear in the next sorted order.

Think about traversal in-order. Who would be the successor of $x$?
The parent of $x$!
Scenario III: Node x Has No Right Subtree and Is Not a Left-Child of an Immediate Parent

Keep moving up the tree until you find a parent which branches from the left().

Stated in Pseudo code.

\[
\begin{align*}
  y &:= \text{parent}(x); \\
  \textbf{while} \ y \neq \text{NULL} \text{ and } x = \text{right}(y) \textbf{ do} \ x := y; \\
  &\quad \quad \quad y := \text{parent}(y); 
\end{align*}
\]
Successor Pseudo-Codes

Verify this code with this tree.

Find successor of
3  →  4
9  →  13
13 → 15
18 → 20

Note that parent(root) = NULL

Algorithm Successor\(\langle x \rangle\)
Input: \(x\) is the input node.
1. if right\(\langle x \rangle\) \(\neq\) NULL
2. then return Minimum(right\(\langle x \rangle\));
3. else
4. \(y :=\) parent\(\langle x \rangle\);
5. while \(y \neq\) NULL and \(x =\) right\(\langle y \rangle\)
6. do \(x :=\) \(y\);
7. \(y :=\) parent\(\langle y \rangle\);
8. return \(y\);

COMP2012H (BST)
Problem

- If we use a “doubly linked” tree, finding parent is easy.

```cpp
class Node
{
    int data;
    Node *left;
    Node *right;
    Node *parent;
};
```

- But usually, we implement the tree using only pointers to the left and right node. 😞 So, finding the parent is tricky.

```cpp
class Node
{
    int data;
    Node *left;
    Node *right;
};
```

For this implementation we need to use a Stack.
Use a Stack to Find Successor

**Algorithm** $\text{Successor}(r, x)$

**Input:** $r$ is the root of the tree and $x$ is the node.

1. initialize an empty stack $S$;
2. while $\text{key}(r) \neq \text{key}(x)$
3. do push($S$, $r$);
4. if $\text{key}(x) < \text{key}(r)$
5. then $r := \text{left}(r)$;
6. else $r := \text{right}(r)$;
7. if right($x$) $\neq$ NULL
8. then return $\text{Minimum}(\text{right}(x))$;
9. else
10. if $S$ is empty
11. then return NULL;
12. else
13. $y := \text{pop}(S)$;
14. while $y \neq$ NULL and $x = \text{right}(y)$
15. do $x := y$;
16. if $S$ is empty
17. then $y := \text{NULL}$;
18. else $y := \text{pop}(S)$;
19. return $y$;

**PART I**
Initialize an empty Stack $S$.

Start at the root node, and traverse the tree until we find the node $x$. Push all visited nodes onto the stack.

**PART II**
Once node $x$ is found, find successor using 3 scenarios mentioned before.

Parent nodes are found by popping the stack!
An Example

**Algorithm** $Successor(r, x)$

**Input:** $r$ is the root of the tree and $x$ is the node.

1. initialize an empty stack $S$;
2. while $key(r) \neq key(x)$
   3. do push($S, r$);
   4. if $key(x) < key(r)$
      5. then $r := left(r)$;
      6. else $r := right(r)$;

$Successor(root, 13)$

**Part I**

Traverse tree from root to find 13

order $\rightarrow$ 15, 6, 7, 13

---

Stack s
Example

Successor(root, 13)
Part II
Find Parent (Scenario III)

y = s.pop()
while y! = NULL and x = right(y)
x = y;
if s.isempty()
y = NULL
else
y = s.pop()
loop
return y

x = 13
y = pop() = 7
y = pop() = 6

y = pop() = 15
->Stop right(15) != x
return y as successor!

Stack s
Another Approach

- Observe that:
  - x must be in the left branch of its successor y, because it is smaller in value.
  - To get to x from left(y), we have the case that we always traverse right, i.e., the value is increasing beyond left(y).
  - If we plot the values from y to x against the nodes visited, it is hence of a “V” shape, starting from y, dropping to some low value, and then increasing gradually to x (a value below y).

- Using stack storing the path from the root to x, we hence can detect the right turn in the reverse path simply as follows:
  - Keep popping the stack until the key is higher than the value x. This must be its successor.

```c
while (!s.empty()) {
    y = s.pop();
    if (y > x) {
        return y; // the successor
    }
}
return NULL; // empty stack; successor not found
```
Insertion

- Insert a new key into the binary search tree
- The new key is always inserted as a new leaf
- Example: Insert 13 ...
Insertion: Another Example

- First add 80 into an existing tree
- Then add 35 into it
Inserting into a BST (1/2)

template<class E, class K>
BSTree<E,K>& BSTree<E,K>::Insert(const E& e)
{
    // Insert e if not duplicate.
    BinaryTreeNode<E> *p = this->root, // search pointer
                      *pp = 0; // parent of p

    // find place to insert
    while (p) {
        // examine p->data
        pp = p;
        // move p to a child
        if (e < p->data) p = p->LeftChild;
        else if (e > p->data) p = p->RightChild;
        else throw BadInput(); // duplicate
    }
}
// get a node for e and attach to pp
BinaryTreeNode<E> *r = new BinaryTreeNode<E> (e);

if (root) {
    // tree not empty
    if (e < pp->data) pp->LeftChild = r;
    else pp->RightChild = r;
} else // insertion into empty tree
    root = r;

return *this;
**BST Deletion: Delete Node z from Tree**

### Three cases for deletion

**Case I**
- **Node z is a leaf**
  - Set z parent’s pointer to z to NULL

**Case II**
- **Node z has exactly 1 (left or right) child**
  - Modify appropriate parent(z) to point to z’s child (Parent adoption)
Case III: Node z Has 2 Children

**Step 1.**
Find successor y of ‘z’ (i.e. y = successor(z))

Since z has 2 children, successor is \( y = \text{minimum}(\text{right}(z)) \)

Successor y of z will have no children or only a right-child. Why? Look at the definition of \( \text{minimum(...)} \)

**Step 2.**
Swap keys of z and y.

Now delete node y (which now has value z)!
This deletion is either case I or II.

(deletion of node “z” is always going to be Case I or II)
Special Case:
Deleting the Root with 1 Child Descendant

- Move the root to the child
A Deletion Example

Three possible cases to delete a node $x$ from a BST

1. The node $x$ is a leaf

COMP2012H (BST)
A Deletion Example (Cont.)

2. The node $x$ has one child

![Tree diagram with node $x$ and free storage label]
A Deletion Example (Cont.)

3. \( x \) has two children

i) Replace contents of \( x \) with inorder successor (smallest value in the right subtree)

ii) Delete node pointed to by \( x_{Succ} \) as described for cases 1 and 2
Removing 40 from (a) results in (b) using the smallest element in the right subtree (i.e., the successor)
Another Deletion Example (Cont.)

- Removing 40 from (a) results in (c) using the largest element in the left subtree (i.e., the predecessor)
Another Deletion Example (Cont.)

- Removing 30 from (c), we may replace the element with either 5 (predecessor) or 31 (successor). If we choose 5, then (d) results.
Deletion Code (1/4)

- First Element Search, and then Convert Case III, if any, to Case I or II

```cpp
template<class E, class K>
BSTree<E,K>& BSTree<E,K>::Delete(const K& k, E& e)
{
    // Delete element with key k and put it in e.
    // set p to point to node with key k (to be deleted)
    BinaryTreeNode<E> *p = root, // search pointer
    *pp = 0; // parent of p

    while (p && p->data != k){
        // move to a child of p
        pp = p;
        if (k < p->data) p = p->LeftChild;
        else p = p->RightChild;
    }
}
```
if (!p) throw BadInput(); // no element with key k

e = p->data; // save element to delete

// restructure tree
// handle case when p has two children
if (p->LeftChild && p->RightChild) {
    // two children convert to zero or one child case
    // find predecessor, i.e., the largest element
    in // left subtree of p
    BinaryTreeNode<E> *s = p->LeftChild,
    *ps = p; // parent of s
    while (s->RightChild) {
        // move to larger element
        ps = s;
        s = s->RightChild;
    }
Deletion Code (3/4)

```c
// move from s to p
p->data = s->data;
p = s;  // move/reposition pointers for deletion
pp = ps;
}

// p now has at most one child
// save child pointer to c for adoption
BinaryTreeNode<E> *c;
if (p->LeftChild) c = p->LeftChild;
else c = p->RightChild;  // may be NULL

// deleting p
if (p == root) root = c;  // a special case: delete root
else {
    // is p left or right child of pp?
    if (p == pp->LeftChild) pp->LeftChild = c;  // adoption
    else pp->RightChild = c;
}
```
Deletion Code (4/4)

```c++
delete p;

return *this;
}
```
Implementation: ADT of Binary Search Tree (BST)

- Construct an empty BST
- Determine if BST is empty
- Search BST for given item
- Insert a new item in the BST
  - Need to maintain the BST property
- Delete an item from the BST
  - Need to maintain the BST property
- Traverse the BST
  - Visit each node exactly once
  - The inorder traversal visits the nodes in ascending order
ADT of a BST

AbstractDataType BSTree {

instances
  binary trees, each node has an element with a key field; all keys are distinct; keys in the left subtree of any node are smaller than the key in the node; those in the right subtree are larger.

operations
  Create(): create an empty binary search tree
  Search(k, e): return in e the element/record with key k
                   return false if the operation fails,
                   return true if it succeeds
  Insert(e): insert element e into the search tree
  Delete(k, e): delete the element with key k and return it in e
  Ascend(): output all elements in ascending order of key

}
A Simple Implementation without Inheritance

- tree_codes *(BST.h and treetester.cpp)*

```cpp
template <typename DataType>
class BST
{
  public:
    // ... member functions supporting BST operations

  private:
    /***** Binary node class *****/
    class BinNode
    {
      public:
        DataType data;
        BinNode * left;
        BinNode * right;

        // ... BinNode constructors

    }; // end of class BinNode declaration

typedef BinNode *BinNodePointer;

    // ... Auxiliary/Utility functions supporting member functions

    /***** Data Members *****/
    BinNodePointer myRoot; // the root of the binary search tree

}; // end of class template declaration
```
Another Implementation with Inheritance, function pointers, and exception handling

- **tree2_codes**
  - Binary search tree is derived from binary tree
  - \( E \) is the record, and \( K \) is the key
  - bst.h:

```cpp
template<class E, class K>
class BSTree : public BinaryTree<E> {
public:
    bool Search(const K& k, E& e) const;
    BSTree<E,K>& Insert(const E& e);
    BSTree<E,K>& InsertVisit(const E& e, void(*visit)(E& u));
    BSTree<E,K>& Delete(const K& k, E& e);
    void Ascend() {InOutput();}
};
```

COMP2012H (BST)
Skeleton of tree2_codes

- **btnode.h**: the node structure to be used in a binary tree

```cpp
template <class T>
class BinaryTreeNode {
    // ... friend functions
    public:
    // ... constructors
    private:
        T data; // data is a record
        BinaryTreeNode<T> *LeftChild, // left subtree
                         *RightChild; // right subtree
};
```

- **binary.h**: binary tree

```cpp
template<class T>
class BinaryTree {
    // ... some friend functions
    public:
    // ... member functions and note the use of
    // function pointers
    private:
        BinaryTreeNode<T> *root; // pointer to root
    //helper/utility functions and static functions
};
```
Code Implementation (tree2_codes)

bst.h

```cpp
#include "bst.h"

#include <iostream> // for std::cout

// May be replaced by recursive codes

template<class E, class K>
bool BSTree<E,K>::Search(const K& k, E &e) const
{// Search for element that matches k.
  // pointer p starts at the root and moves through
  // the tree looking for an element with key k
  BinaryTreeNode<E> *p = this->root;
  while (p) // examine p->data
    if (k < p->data) p = p->LeftChild; // implicit cast
    else if (k > p->data) p = p->RightChild;
    else {// found element
      e = p->data; // copy the record to e
      return true;}
  return false;
}
```

datatype.h: DataType is to be used in the binary node with field data

```cpp
#ifndef DataType_
#define DataType_

class DataType {
  friend ostream& operator<<(ostream&, DataType);
  public:
    operator int() const {return key;} // implicit cast to obtain key
    int key; // element key, maybe hashed from ID
    char ID; // element identifier
  }

  ostream& operator<<(ostream& out, DataType x)
  { out << x.key << ' ' << x.ID << ' '; return out; }

#endif
```
Time Complexity of Binary Search Trees

- Find(x) \( O(\text{height of tree}) \)
- Min(x) \( O(\text{height of tree}) \)
- Max(x) \( O(\text{height of tree}) \)
- Insert(x) \( O(\text{height of tree}) \)
- Delete(x) \( O(\text{height of tree}) \)
- Traverse \( O(N) \)
Binary Search Trees

- **Problem**
  - How can we predict the height of the tree?

- Many trees of different shapes can be composed of the same data

- How to control the tree shape?
Problem of Lopsidedness

- Trees can be unbalanced
- Not all nodes have exactly 2 child nodes
Problem of Lopsidedness

- Trees can be totally lopsided
- Suppose each node has a right child only
- Degenerates into a linked list

Processing time affected by "shape" of tree
Binary Search Tree

(we say this tree is “balanced”)

Tree 1
Same data as Tree 2

Tree 2
Same data as Tree 1

Which tree would you prefer to use?
Tree Examples

(Tree resulting from randomly generated input)
Tree Examples

(Unbalanced tree)
How Fast is Sorting Using BST?

- $n$ numbers ($n$ large) are to be sorted by first constructing a binary tree and then read them in inorder manner
- **Bad case:** the input is more or less sorted
  - A rather “linear” tree is constructed
  - Total steps in constructing a binary tree: $1 + 2 + \ldots + n = n(n+1)/2 \sim n^2$
  - Total steps in traversing the tree: $n$
  - Total $\sim n^2$
- **Best case:** the binary tree is constructed in a balanced manner
  - Depth after adding $i$ numbers: $\log(i)$
  - Total steps in constructing a binary tree: $\log 1 + \log 2 + \log 3 + \log 4 + \ldots + \log n < \log n + \log n + \ldots + \log n = n \log n$
  - Total steps in traversing the tree: $n$
  - Total $\sim n \log n$, much faster
- It turns out that one cannot sort $n$ numbers faster than $n \log n$
- For any arbitrary input, one can indeed construct a rather balanced binary tree with some extra steps in insertion and deletion
  - E.g., An AVL tree
An AVL Tree \rightarrow A Rather Balanced Tree for Efficient BST Operations (See Animation)

(Balanced Tree . . This is actually a very good tree called AVL tree)