Binary Trees

- **Linked lists: efficient insertion/deletion, inefficient search**
  - Vector: search can be efficient, insertion/deletion not

- **Binary trees: efficient insertion, deletion, and search**
  - Trees used in many contexts, not just for searching, e.g., expression trees
  - Search in $O(\log n)$ like sorted array
  - Insertion/deletion $O(1)$ like list, *once location found!*
  - Binary trees are inherently recursive, difficult to process trees non-recursively, but possible
    - Recursion never required, often makes coding simpler
From doubly-linked lists to binary trees

- Instead of using prev and next to point to a linear arrangement, use them to divide the universe in half
  - Similar to binary search, everything less goes left, everything greater goes right

- How do we search?
- How do we insert?
See code in treeunique.cpp

class DoubleListUnique
{
    protected:
    struct Node {
        string info;
        Node * next;
        Node * prev;
        Node (const string& s,
               Node * after, Node * before)
            : info(s),
              next(after),
              prev(before)
        {
        }
        virtual Node * addHelper(const string& s,
                                  Node * list);
    };
};
Helper function in treeunique.cpp

- How is this helper different from DoubleListUnique helper?

```cpp
Node * addHelper(const string& s, Node * list) {
    if (list == 0) {
        myCount++;
        return new Node(s, 0, 0);
    }
    if (list->info == s) return list;
    if (list->info < s) {
        list->next = addHelper(s, list->next);
    } else {
        list->prev = addHelper(s, list->prev);
    }
    return list;
}
```
Basic tree definitions

- **Binary tree is a structure:**
  - empty
  - *root node with left and right subtrees*

- **Terminology:** parent, children, leaf node, internal node, depth, height, path
  - link from node N to M then N is *parent* of M
    - M is *child* of N
  - *leaf node* has no children
    - internal node has 1 or 2 children
  - *path* is sequence of nodes, N₁, N₂, ..., Nₖ
    - Nᵢ is parent of Nᵢ₊₁
    - sometimes edge instead of node
  - *depth* (level) of node: length of root-to-node path
    - level of root is 1 (measured in nodes)
  - *height* of node: length of longest node-to-leaf path
    - height of tree is height of root
Printing a search tree in order

• **When is root printed?**
  - After left *subtree*, before right *subtree*.

```c
void visit(Node * t)
{
    if (t != 0) {
        visit(t->left);
        cout << t->info << endl;
        visit(t->right);
    }
}
```

• **Inorder traversal**
Insertion and Find? Complexity?

- How do we search for a value in a tree, starting at root?
  - Can do this both iteratively and recursively, contrast to printing which is very difficult to do iteratively
  - How is insertion similar to search?

- What is complexity of print? Of insertion?
  - Is there a worst case for trees?
  - Do we use best case? Worst case? Average case?

- How do we define worst and average cases
  - For trees? For vectors? For linked lists? For vectors of linked-lists?
Implementing binary trees

- Trees can have many shapes: short/bushy, long/stringy
  - if height is $h$, number of nodes is between $h$ and $2^h - 1$
  - single node tree: height = 1, if height = 3

- C++ implementation, similar to doubly-linked list

```c++
struct Tree
{
    string info;
    Tree * left;
    Tree * right;
    Tree(const string& s, Tree * lptr, Tree * rptr)
        : info(s), left(lptr), right(rptr)
            { }
    
};
```
Tree functions

- **Compute height of a tree, what is complexity?**

  ```cpp
  int height(Tree * root) {
    if (root == 0) return 0;
    else {
      return 1 + max(height(root->left), height(root->right));
    }
  }
  ```

- **Modify function to compute number of nodes in a tree, does complexity change?**
  - What about computing number of leaf nodes?
Tree traversals

- **Different traversals useful in different contexts**
  - Inorder prints search tree in order
    - Visit left-subtree, process root, visit right-subtree
  - Preorder useful for reading/writing trees
    - Process root, visit left-subtree, visit right-subtree
  - Postorder useful for destroying trees
    - Visit left-subtree, visit right-subtree, process root
Insertion into search tree

• Simple recursive insertion into tree

```cpp
void insert(Tree * & t, const string & s)
// pre: t is a search tree
// post: s inserted into t, t is a search tree
{
    if (t == 0)            t = new Tree(s, 0, 0);
    else if (s <= t->left) insert(t->left, s);
    else                   insert(t->right, s);
}
```

• Note: in each recursive call, the parameter t in the called clone is either the left or right pointer of some node in the original tree
  ➢ Why is this important?
  ➢ Why must t be passed by reference?
  ➢ For alternatives see treeunique.cpp
Balanced Trees and Complexity

- **A tree is height-balanced if**
  - Left and right subtrees are height-balanced
  - Left and right heights differ by at most one

```c
bool isBalanced(Tree * root) {
    if (root == 0) return true;
    else {
        return isBalanced(root->left) &&
               isBalanced(root->right) &&
               abs(height(root->left) - height(root->right)) <= 1;
    }
}
```
What is complexity?

- Assume trees are “balanced” in analyzing complexity
  - Roughly half the nodes in each subtree
  - Leads to easier analysis

- How to develop recurrence relation?
  - What is $T(n)$?
  - What other work is done?

- How to solve recurrence relation
  - Plug, expand, plug, expand, find pattern
  - A real proof requires induction to verify correctness
Danny Hillis

- The third culture consists of those scientists and other thinkers in the empirical world who, through their work and expository writing, are taking the place of the traditional intellectual in rendering visible the deeper meanings of our lives, redefining who and what we are.

*(Wired 1998)* And now we are beginning to depend on computers to help us evolve new computers that let us produce things of much greater complexity. Yet we don't quite understand the process - it's getting ahead of us. We're now using programs to make much faster computers so the process can run much faster.

That's what's so confusing - technologies are feeding back on themselves; we're taking off. We're at that point analogous to when single-celled organisms were turning into multicelled organisms. We are amoebas and we can't figure out what the hell this thing is that we're creating.