Data Structures and Algorithms

Roberto Sebastiani
roberto.sebastiani@disi.unitn.it
http://www.disi.unitn.it/~rseba

- Week 06 -

B.S. In Applied Computer Science
Free University of Bozen/Bolzano
academic year 2010-2011

Acknowledgements & Copyright Notice

Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

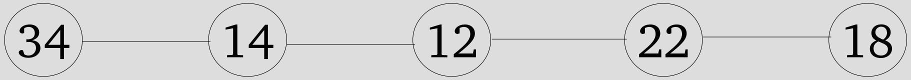
Dictionaries

- A *dictionary D* is a dynamic data structure with operations:
  - **Search(D, k)** – *returns a pointer x to an element such that x.key = k (null otherwise)*
  - **Insert(D, x)** – *adds the element pointed to by x to D*
  - **Delete(D, x)** – *removes the element pointed to by x from D*
- An element has a *key* and *data* part.

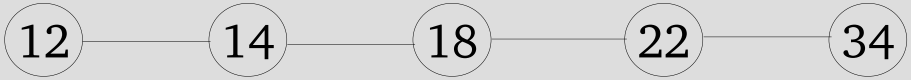
Ordered Dictionaries

- In addition to dictionary functionality, we may want to support operations:
  - **Min(D)**
  - **Max(D)**
- and
  - **Predecessor(D, k)**
  - **Successor(D, k)**
- These operations require keys that are *comparable (ordered domain)*.

A List-Based Implementation

- Unordered list 

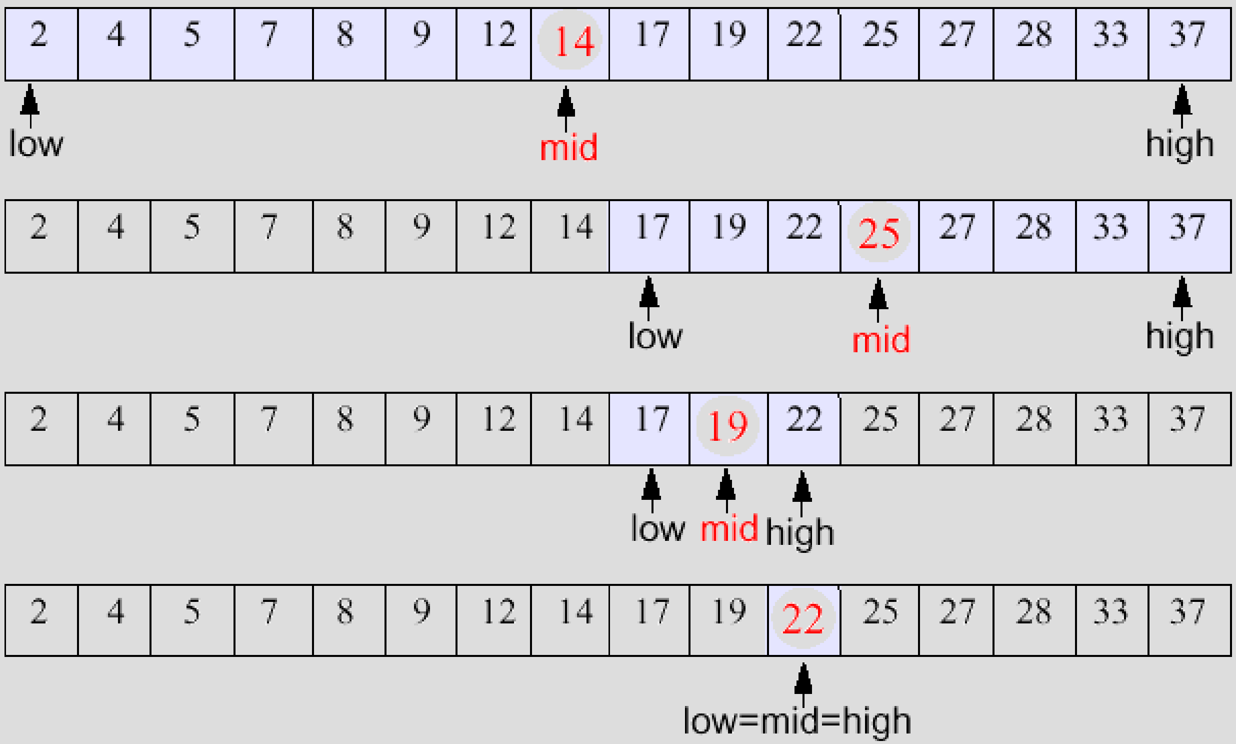
  - search, min, max, predecessor, successor: *O(n)*
  - insertion, deletion: *O(1)*

- Ordered list 

  - search, insert, delete: *O(n)*
  - min, max, predecessor, successor: *O(1)*

Refresher: Binary Search

- Narrow down the search range in stages
  - findElement(22)

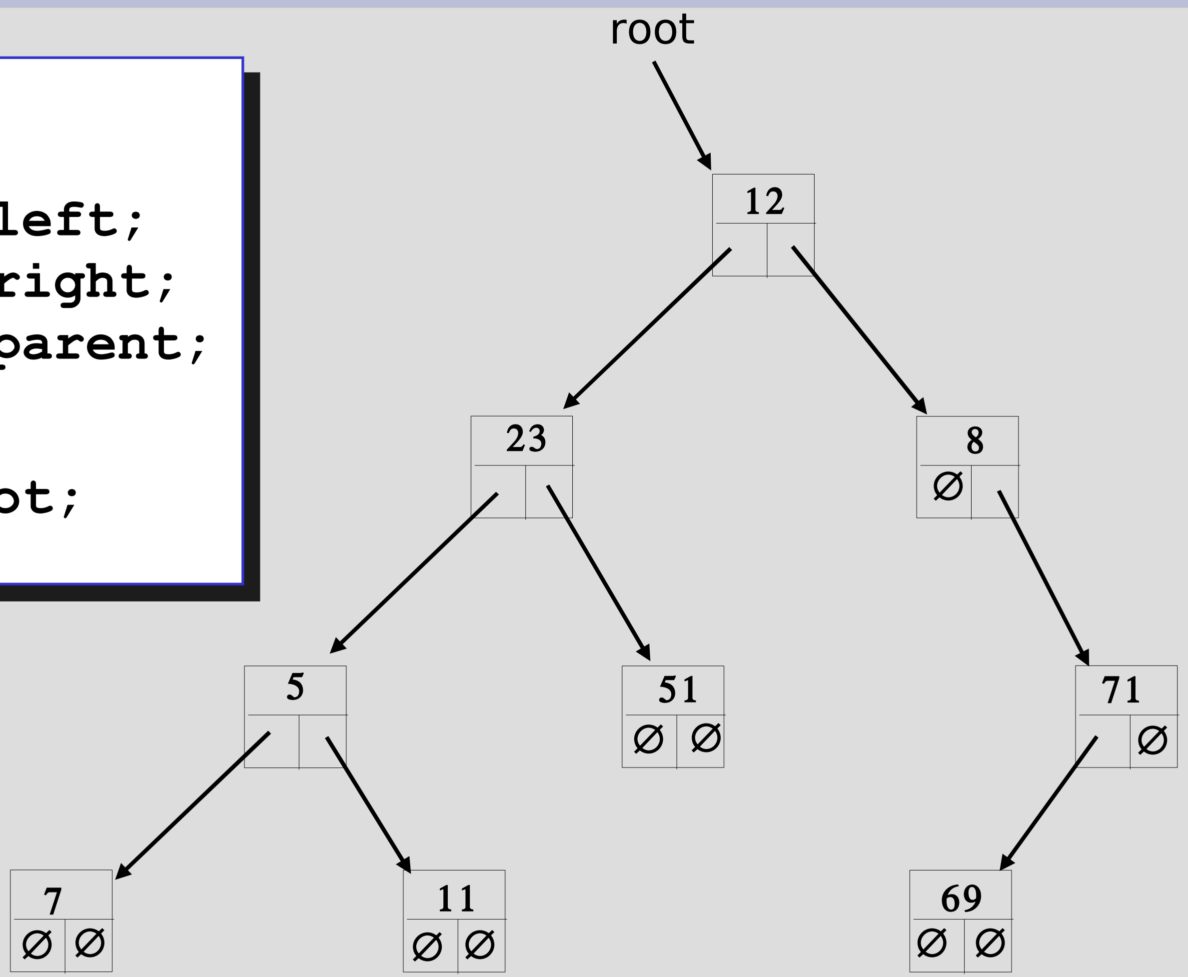


Run Time of Binary Search

- The range of candidate items to be searched is halved after comparing the key with the middle element.
- Binary search runs in $O(\log n)$ time.
- What about insertion and deletion?
  - search: $O(\log n)$
  - insert, delete: $O(n)$
  - min, max, predecessor, successor: $O(1)$
- The idea of a binary search can be extended to dynamic data structures ➔ binary trees.

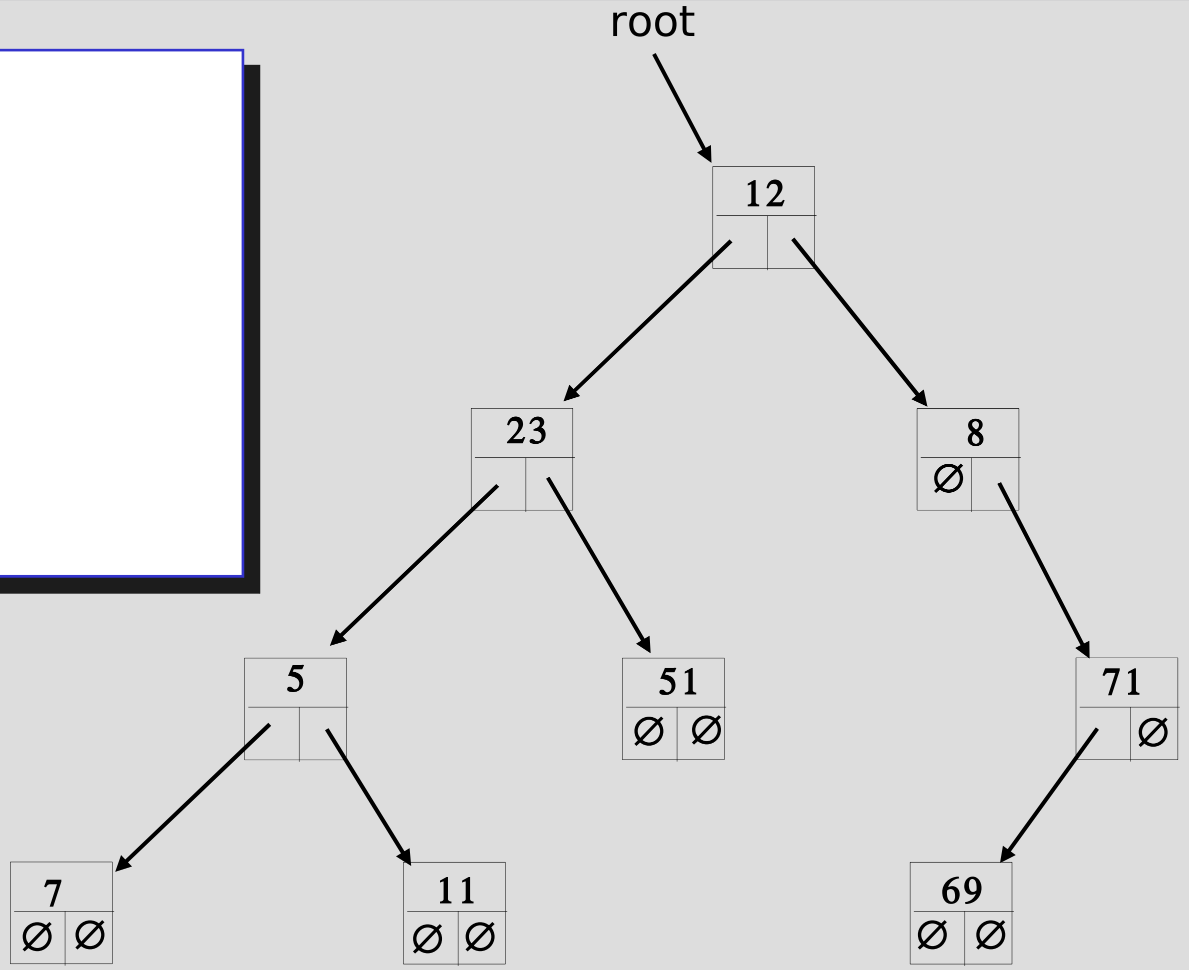
Binary Tree ADT (C)

```
struct node {
  int val;
  struct node* left;
  struct node* right;
  struct node* parent;
}

struct node* root;
```



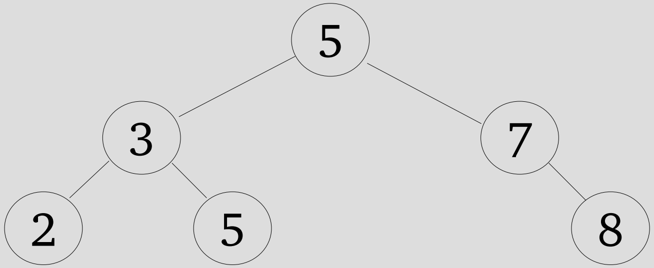
Binary Tree ADT (C)

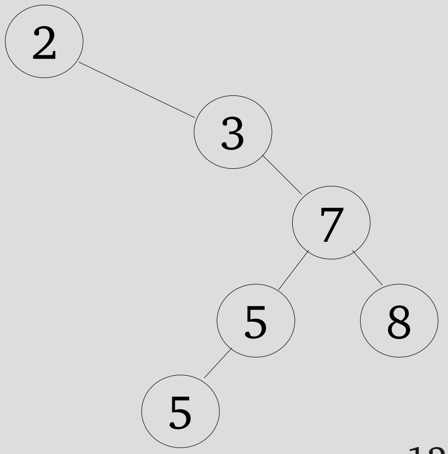
```
class node {
  int val;
  node left;
  node right;
  node parent;
}

node root;
```



Binary Search Trees

- A **binary search tree** (BST) is a binary tree T with the following properties:
  - each internal node stores an item *(k,e)* of a dictionary
  - keys stored at nodes in the **left subtree** of *v* are **less than or equal** to *k*
  - keys stored at nodes in the **right subtree** of *v* **are greater than or equal** to *k*
- Example BSTs for 2, 3, 5, 5, 7, 8





Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Tree Walks

- Keys in a BST can be printed using "tree walks"
- Keys of each node printed between keys in the left and right subtree – *inorder* tree traversal

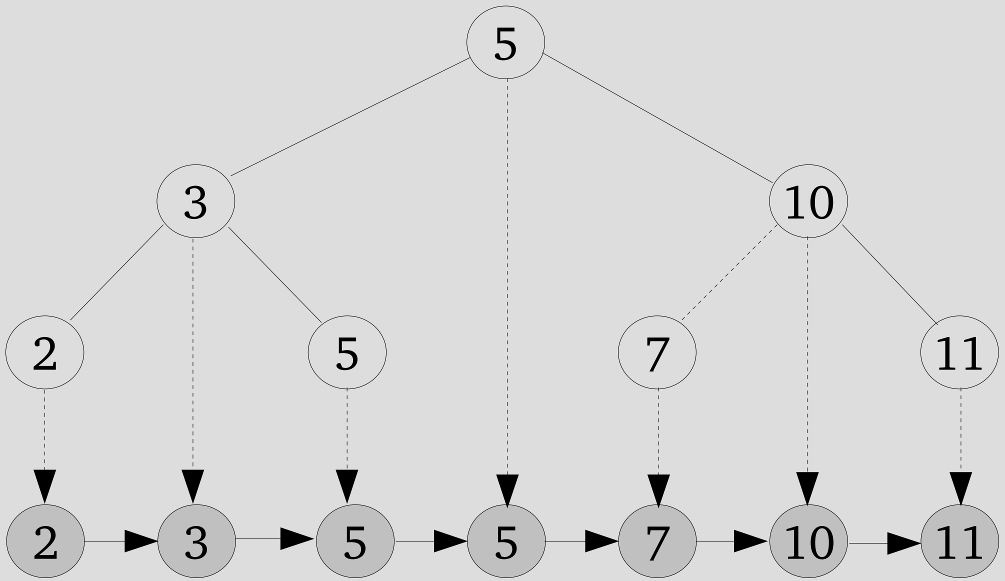
```
InorderTreeWalk(x)
  if x ≠ NIL then
     InorderTreeWalk(x.left)
     print x.key
     InorderTreeWalk(x.right)
```

Tree Walks/2

- InorderTreeWalk is a divide-and-conquer algorithm.
- It prints all elements in monotonically increasing order.
- Running time $\Theta(n)$.

Tree Walks/2

- **Inorder tree walk** can be thought of as a projection of the BST nodes onto a one dimensional interval.



Tree Walks/3

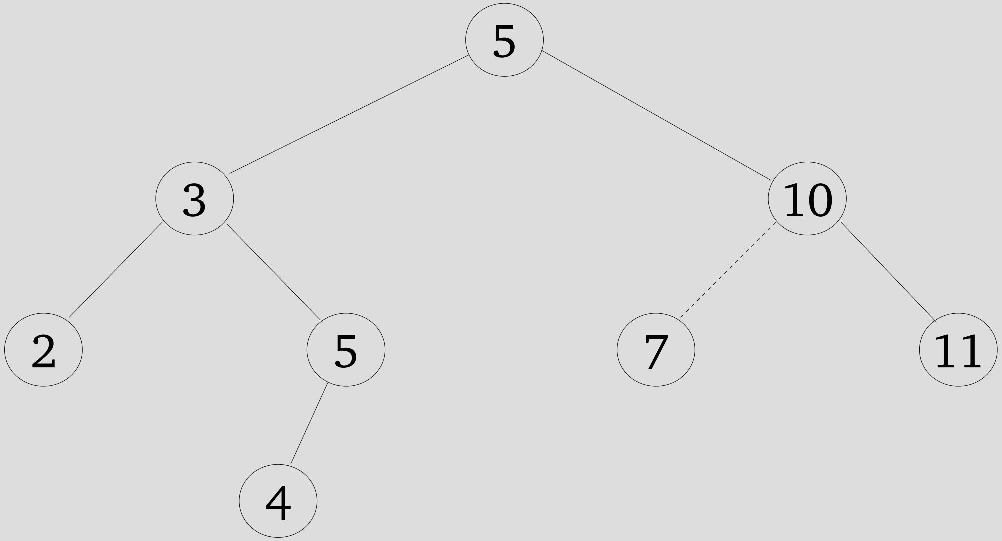
Other forms of tree walk:

- A **preorder tree walk** processes each node before processing its children.
- A **postorder tree walk** processes each node after processing its children.

Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Searching a BST



- To find an element with key k in a tree T
  - compare k with $T.key$
  - if $k < T.key$, search for k in $T.left$
  - otherwise, search for k in $T.right$

Pseudocode for BST Search

- Recursive version: divide-and-conquer

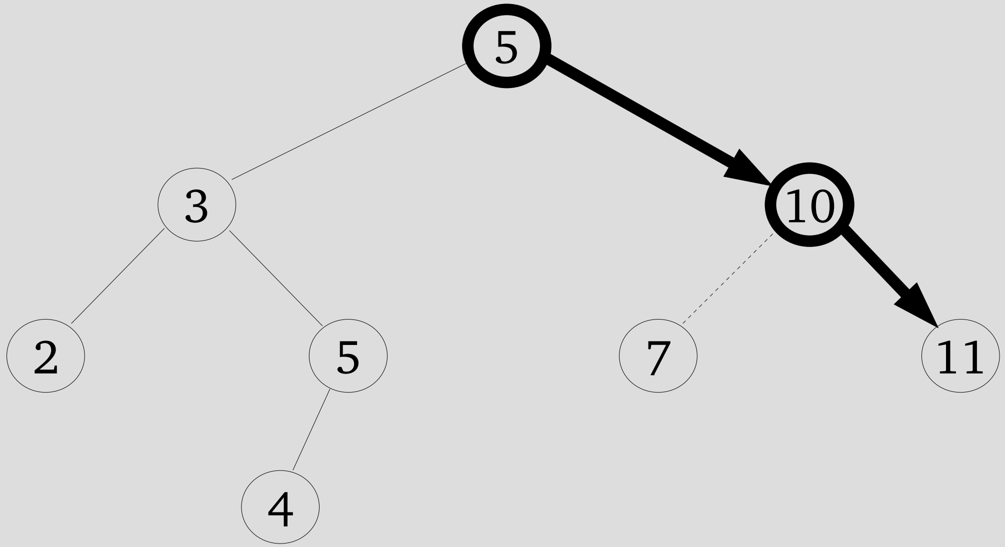
```
Search(T,k)
if T = NIL then return NIL
if k = T.key then return T
if k < T.key
  then return Search(T.left,k)
  else return Search(T.right,k)
```

- Iterative version

```
Search(T,k)
x := T
while x ≠ NIL and k ≠ x.key do
  if k < x.key
    then x := x.left
    else x := x.right
return x
```

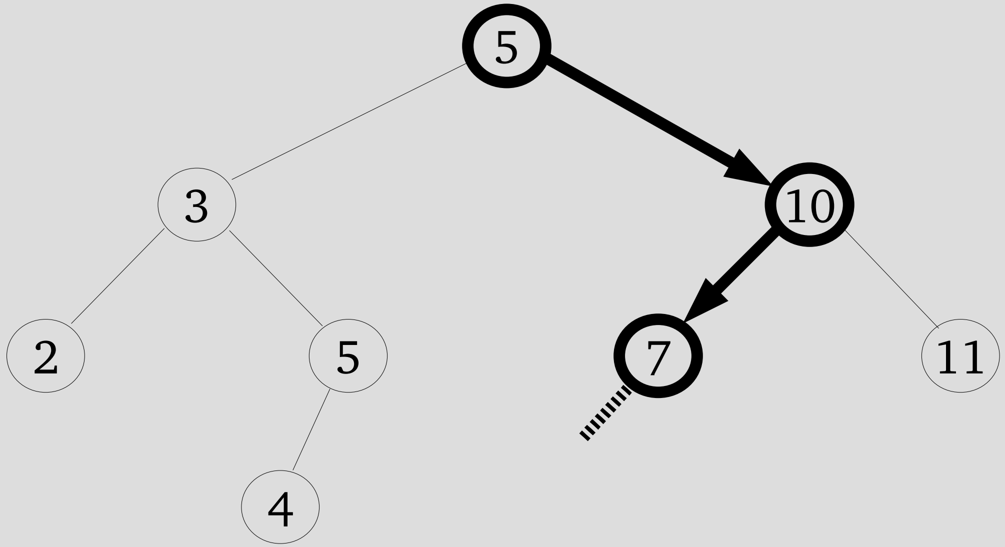
Search Examples

- Search(T, 11)



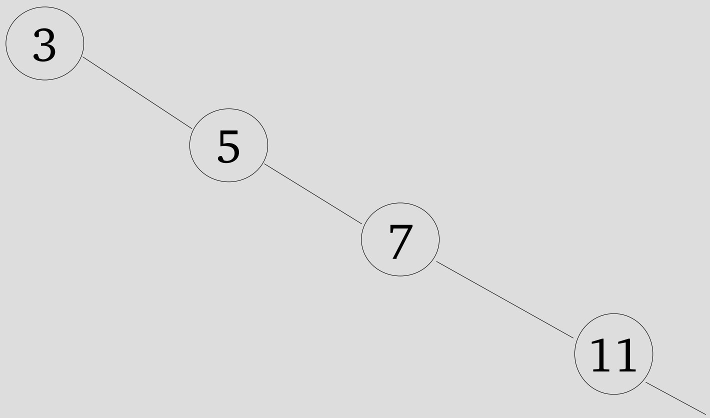
Search Examples/2

- Search(T, 6)



Analysis of Search

- Running time on tree of height h is $O(h)$
- After the insertion of n keys, the worst-case running time of searching is $O(n)$



BST Minimum (Maximum)

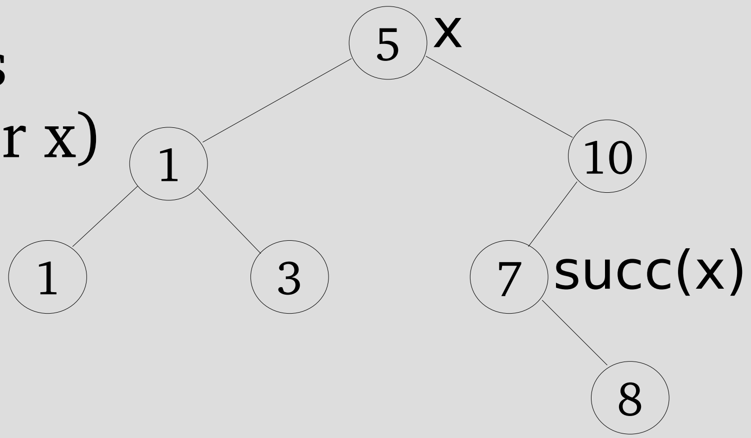
- Find the minimum key in a tree rooted at x.

```
TreeMinimum(x)
while x.left ≠ NIL do
   x := x.left
return x
```

- Maximum: same, x.right instead of x.left
- Running time *O(h)*, i.e., it is proportional to the height of the tree.

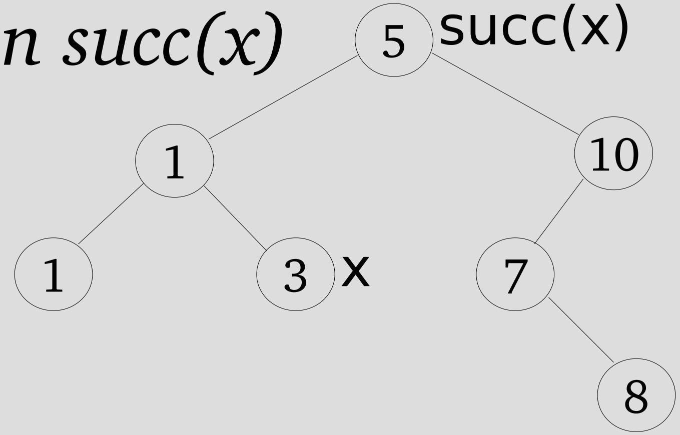
Successor

- Given *x*, find the node with the smallest key greater than *x*.key.
- We can distinguish two cases, depending on the right subtree of *x*
- Case 1: The right subtree of x is non-empty (succ(x) inserted after x)
  - successor is the leftmost node in the right subtree.
  - this can be done by returning TreeMinimum(x.right).



Successor/2

- Case 2: the right subtree of *x* is empty (succ(x), if any, was inserted before x).
  - The successor (if any) is the lowest ancestor of *x* whose left subtree contains *x*.
  - *Note: it x had a right child, then it would be smaller than succ(x)*



Successor Pseudocode

```
TreeSuccessor(x)
if x.right ≠ NIL
  then return TreeMinimum(x.right)
y := x.parent
while y ≠ NIL and x = y.right
  x := y
  y := y.parent
return y
```

- For a tree of height *h*, the running time is *O(h)*.
- *Note: no comparison among keys needed!*

Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

BST Insertion

- The basic idea derives from searching:
  - construct an element p whose left and right children are NULL and insert it into T
  - find location in T where p belongs to (as if searching for $p.key$),
  - add p there
- The running time on a tree of height h is $O(h)$.

BST Insertion Code (C)

- Have a ”one step delayed” pointer.

```
struct node* insert(struct node* p, struct node* r) {
  struct node* y = NULL; struct node* x = r;
  while (x != NULL) {
    y := x;
    if (x->key < p->key) x = x->right;
    else x = x->left;
  }
  if (y == NULL) {r = p;p->partent=null}
  else if (y->key < p->key) y->right = p;
  else y->left = p;
  p->parent = u;
  return r;
}
```

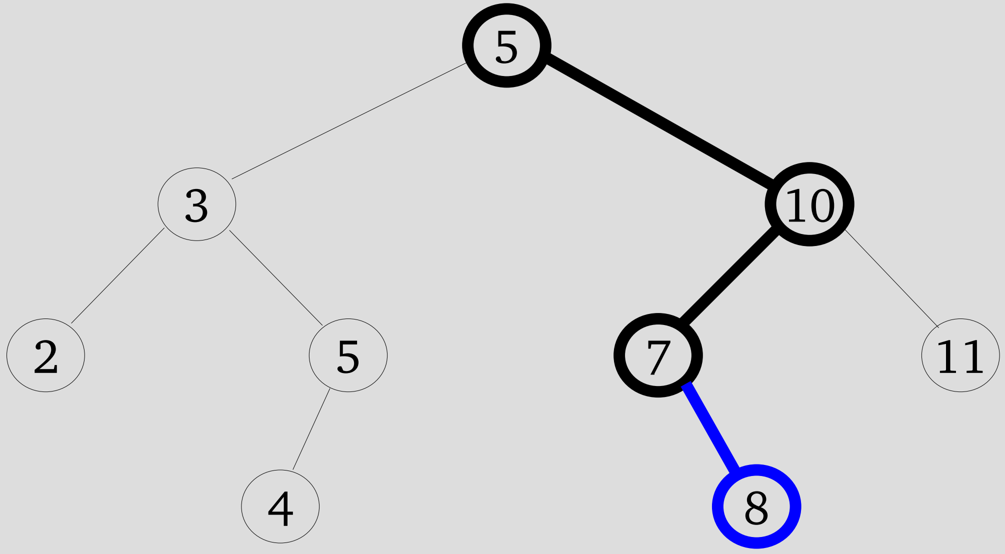
BST Insertion Code (java)

- Have a ”one step delayed” pointer.

```
node insert(node p, node r) { //insert p in r
   node y = NULL; node x = r;
   while (x != NULL) {
      y := x;
      if (x.key < p.key) x = x.right;
      else x = x.left;
   }
   if (y == NULL) {r = p; p.parent=null;}// r is empty
   else if (y.key < p.key) y.right = p;
   else y.left = p;
   p.parent =y;
   return r;
}
```

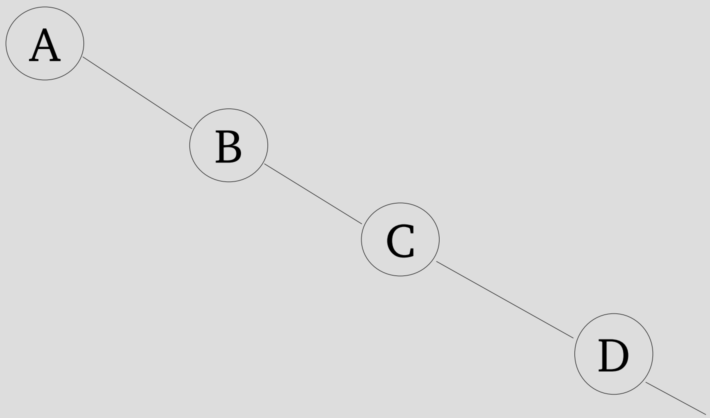
BST Insertion Example

- Insert 8



BST Insertion: Worst Case

- In what kind of sequence should the insertions be made to produce a BST of height *n*?



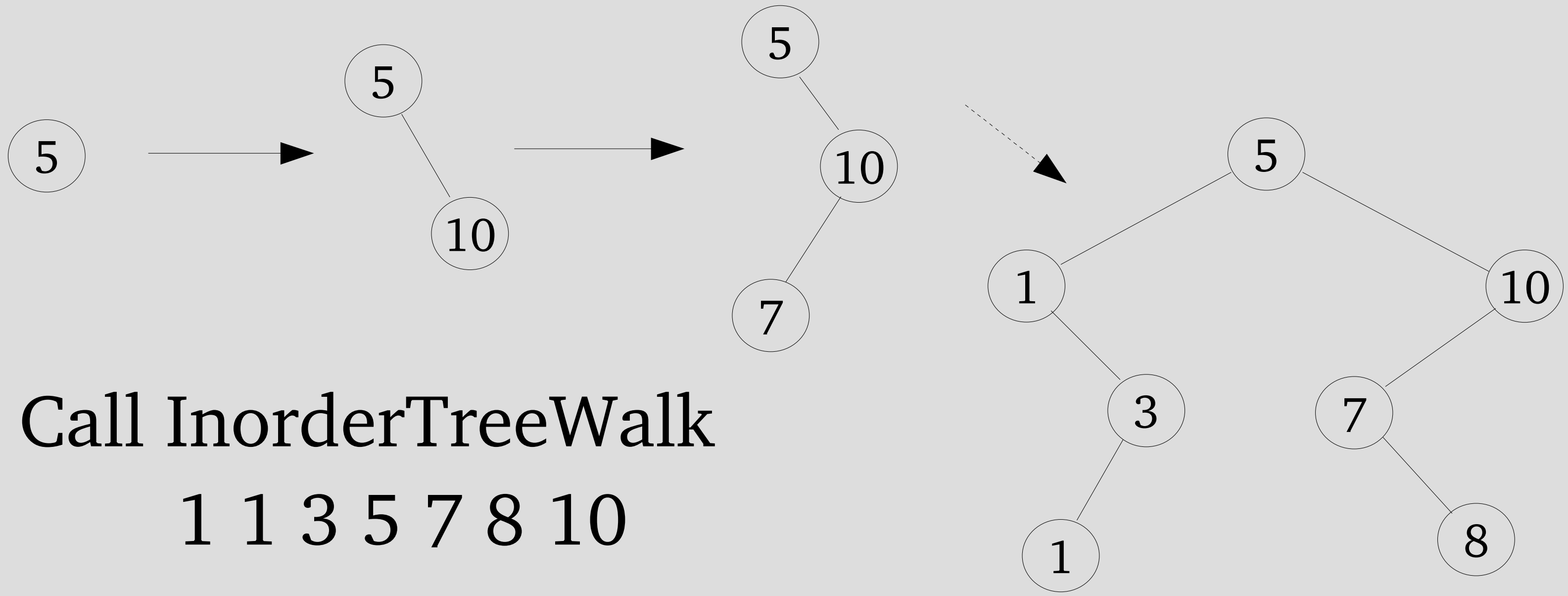
BST Sorting

- Use TreeInsert and InorderTreeWalk to sort a list of n elements, A

```
TreeSort(A)
T := NIL
for i := 1 to n
  TreeInsert(T, BinTree(A[i]))
InorderTreeWalk(T)
```

BST Sorting/2

- Sort the following numbers
  5 10 7 1 3 1 8
- Build a binary search tree



- Call InorderTreeWalk

  1 1 3 5 7 8 10

Data Structures and Algorithms
Week 6

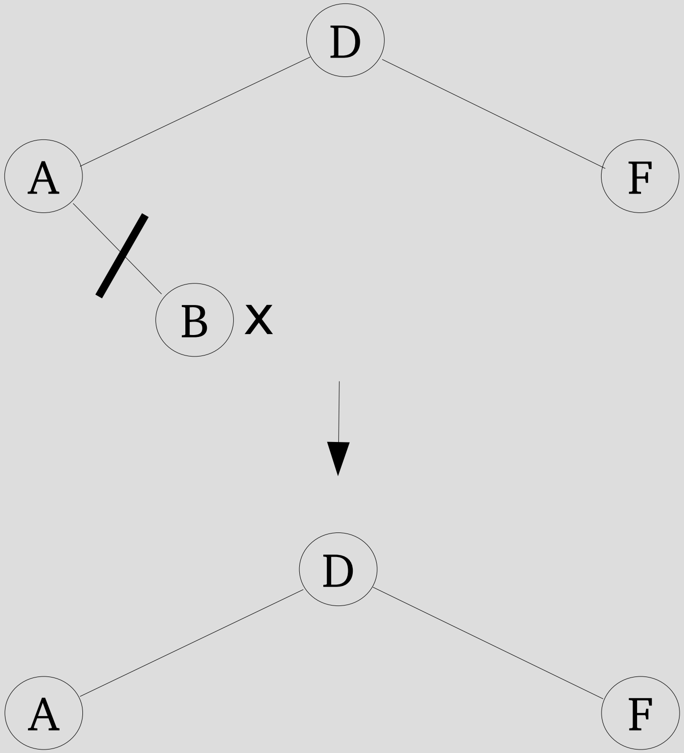
- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Deletion

- Delete node x from a tree T
- We can distinguish three cases
  - x has no child
  - x has one child
  - x has two children

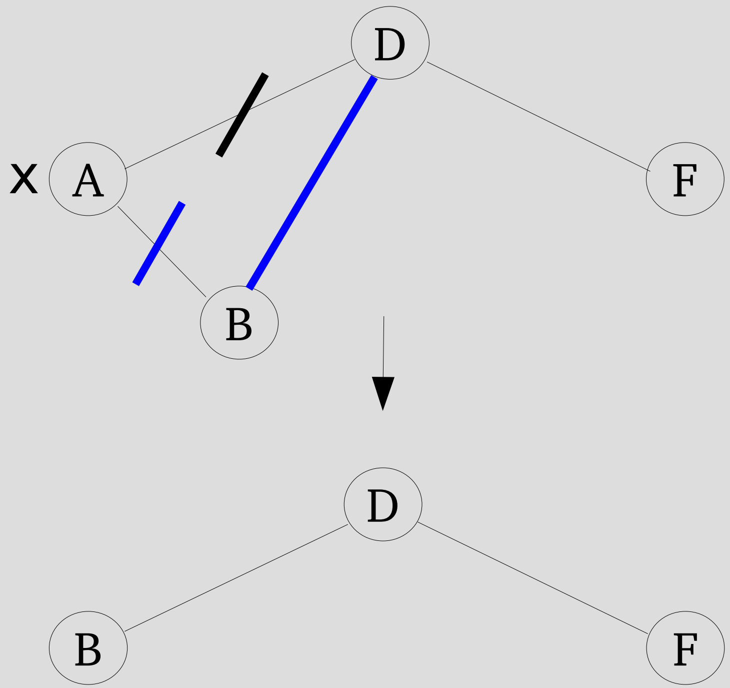
Deletion Case 1

- If *x* has no children: simply remove *x*



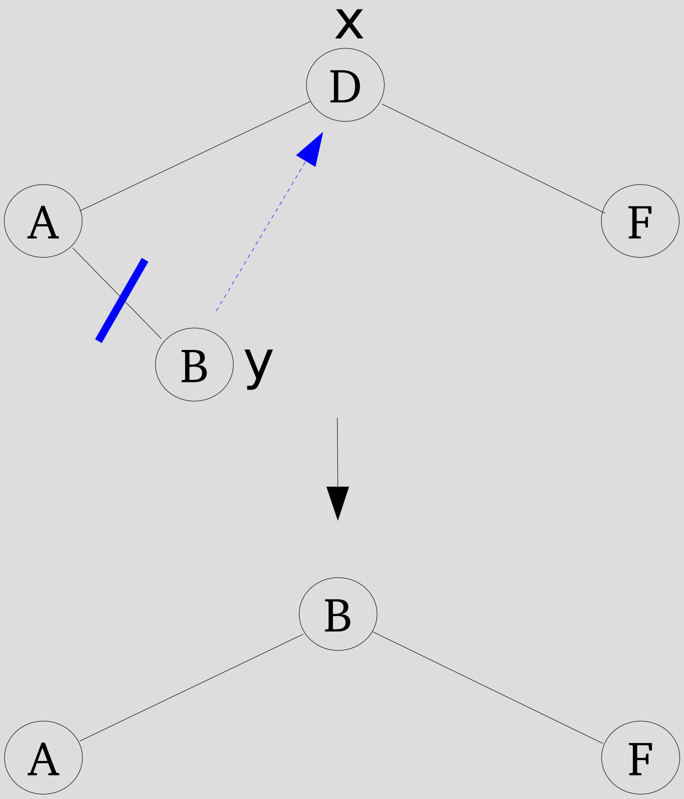
Deletion Case 2

- If *x* has exactly one child, make parent of x point to that child and delete x.



Deletion Case 3

- If *x* has two children:
  - find the largest child y in the left subtree of x (i.e. y is predecessor(x))
  - Recursively remove y (note that y has at most one child), and
  - replace x with y.
- “Specular” version with successor(x) (CLRS)



BST Deletion Code (C)

- Version without “parent” field

```
struct node* delete(struct node* root,
                    struct node* x) {

  u = root; v = NULL;
  while (u != x) {
    v := u;
    if (x->key < u->key) u := u->left;
    else u := u->right;
  } // v points to a parent of x (if any)

  …
```

BST Deletion Code (C)/2

- x has less than 2 children
- Fix pointer of parent of x

```
...
  if (u->right == NULL) {
    if (v == NULL) root = u->left;
    else if (v->left == u) v->left = u->left;
    else v->right = u->left;
  else if (u->left == NULL) {
    if (v == NULL) root = u->right;
    else if (v->left == u) v->left = u->right;
    else v->right = u->right;
  else {
...
```

BST Deletion Code (C)/3

- x has 2 children

```
p = x->left; q = p;
while (p->right != NULL) { q:=p; p:=p->right; }

if (v == NULL) root = p;
else if (v->left == u) v->left = p;
else v->right = p;

p->right = u->right;
if (q != p) {
  q->right = p->left;
  p->left = u->left;
}
return root
```

BST Deletion Code (java)

- Version without "parent" field

```
node delete(node root, node x) {

  u = root; v = NULL;
  while (u != x) {
    v := u;
    if (x.key < u.key) u := u.left;
    else u := u.right;
  } // v points to a parent of x (if any)

  …
```

BST Deletion Code (java)/2

- x has less than 2 children
- Fix pointer of parent of x

```
…
  if (u.right == NULL) {
    if (v == NULL) {root=u.left;
    else if (v.left == u) v.left = u.left;
    else v.right = u.left;
  else if (u.left == NULL) {
    if (v == NULL) root = u.right;
    else if (v.left == u) v.left = u.right;
    else v.right = u.right;
  else {
…
```

BST Deletion Code (java)/3

- x has 2 children

```
p = x.left; q = p;
while (p.right != NULL) { q:=p; p:=p.right; }

if (v == NULL) root = p;
else if (v.left == u) v.left = p;
else v.right = p;

p.right = u.right;
if (q != p) {
  q.right = p.left;
  p.left = u.left;
}
return root
```

BST Deletion Code (java)

- Version with “parent” field

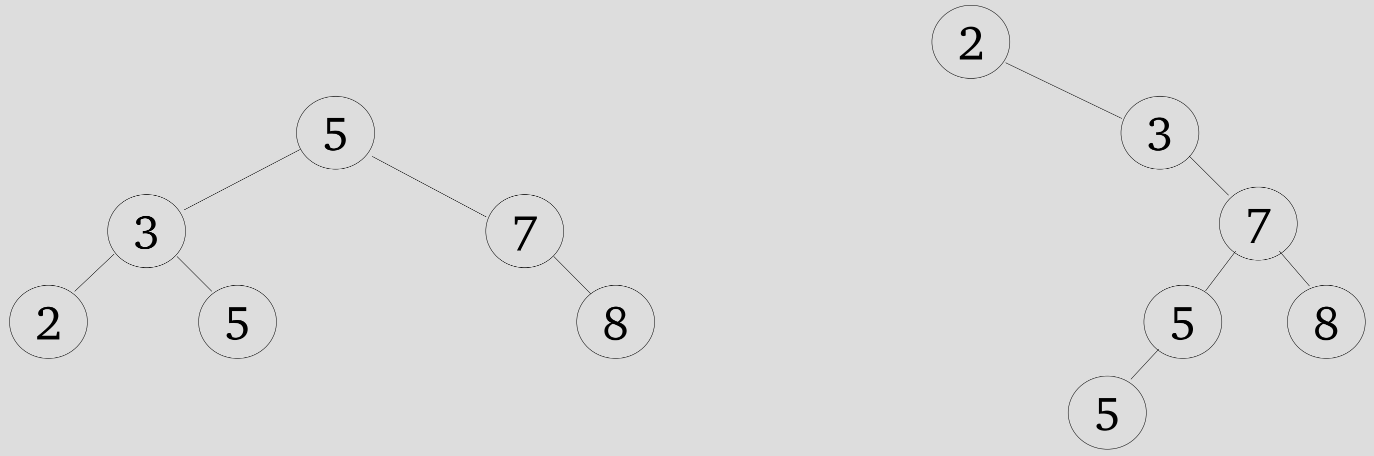
```
node delete(node root, key v) {
  node t; //the node in the tree whose key is v
  node s; //the node which will be deleted
  node r; //the child of s
  t = search(root,v);
  if (t==null) return;
  if (t.l==null||t.r == null) // v has 0,1 children
     s = t;
  else // v has 2 children
     s = succ(root,v); //other version with pred()
  // now s has at most one child
  …
```

BST Deletion Code (java)

```
...
// now s has at most one child
   if (s.l != null) r = s.l;
    else r = s.r;
   if (r!=null) r.p = s.p;
   if (s.p == null)  // s is the root
       root = r;
   else if (s == s.p.l) // s is a left child
       s.p.l = r;
   else // s is a right child
       s.p.r = r;
   if (s != t) // v had 2 children
       t.k = s.k;
 }
```

Balanced Binary Search Trees

- Problem: execution time for tree operations is Θ(h), which in worst case is Θ(n).
- Solution: balanced search trees *guarantee* small height $h = O(\log n)$.



Suggested exercises

- Implement a binary search tree with the following functionalities:
  - init, max, min, successor, predecessor, search (iterative & recursive), insert, delete (both swap with succ and pred), print, print in reverse order
  - TreeSort

Suggested exercises/2

Using paper & pencil:

- draw the trees after each of the following operations, starting from an empty tree:
  1. Insert 9,5,3,7,2,4,6,8,13,11,15,10,12,16,14
  2. Delete 16, 15, 5, 7, 9 (both with succ and pred strategies)
- simulate the following operations after 1:
  - Find the max and minimum
  - Find the successor of 9, 8, 6

Data Structures and Algorithms
Week 6

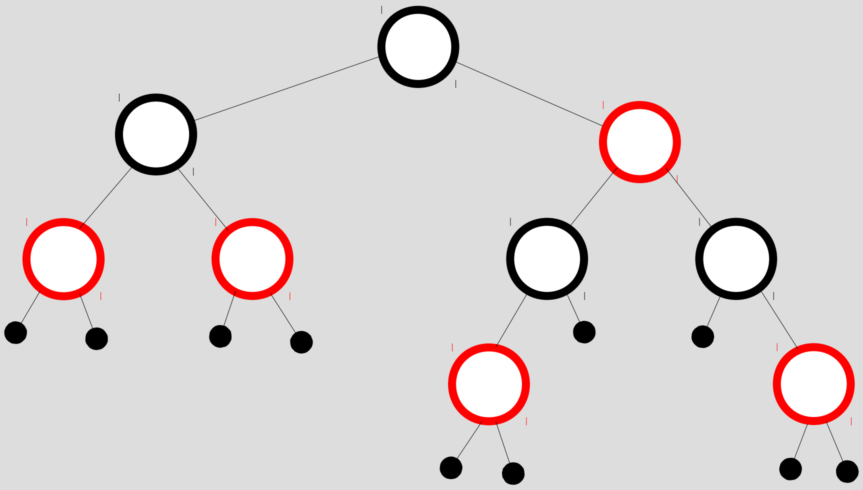
- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Data Structures and Algorithms
Week 6

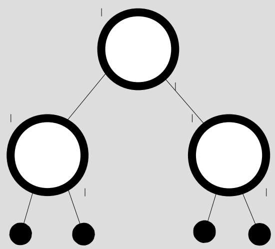
- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Red/Black Trees

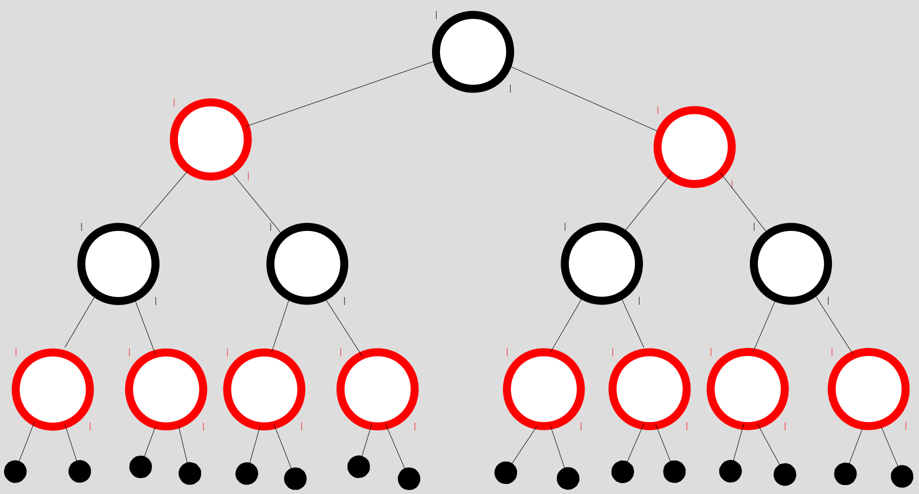
- A **red-black** tree is a binary search tree with the following properties:
  1. Nodes (or incoming edges) are colored **red** or **black**
  2. NULL leaves are **black**
  3. The root is **black**
  4. No two consecutive **red nodes** on any root-leaf path.
  5. Same number of black nodes on any root-leaf path (called *black height* of the tree).



RB-Tree Properties



- Some measures
  - n – # of internal nodes
  - h – height
  - bh – black height
- $2^{bh} - 1 \leq n$
- $bh \geq h/2$
- $2^{h/2} \leq n + 1$
- $h \leq 2 \log(n + 1)$
- BALANCED!



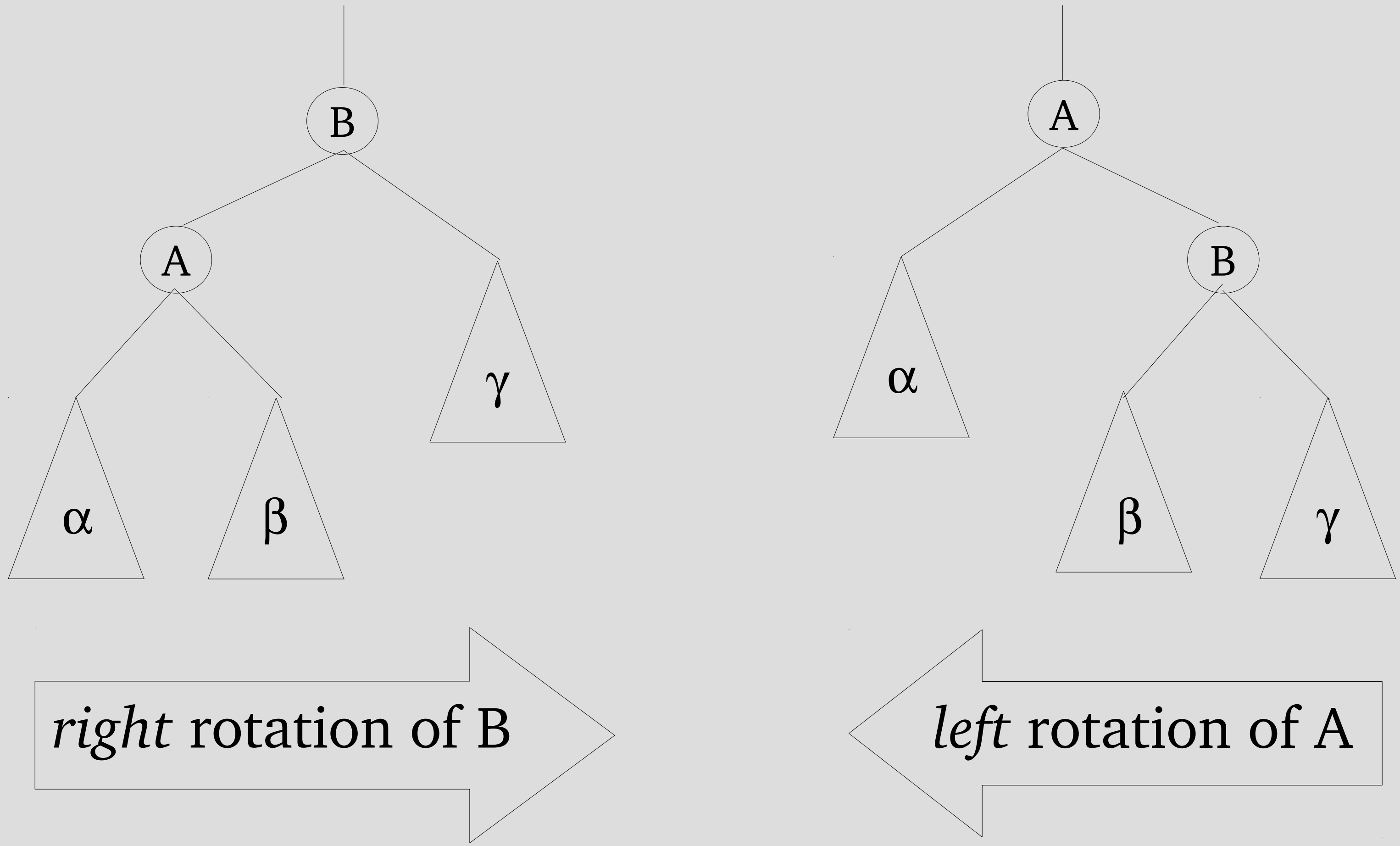
RB-Tree Properties/2

- Operations on a binary-search tree (search, insert, delete, ...) can be accomplished in $O(h)$ time.
- The RB-tree is a binary search tree, whose height is bound by $2 \log(n+1)$, thus the operations run in $O(\log n)$.
  - Provided that we can maintain red-black tree properties spending no more than $O(h)$ time on each insertion or deletion.

Data Structures and Algorithms
Week 6

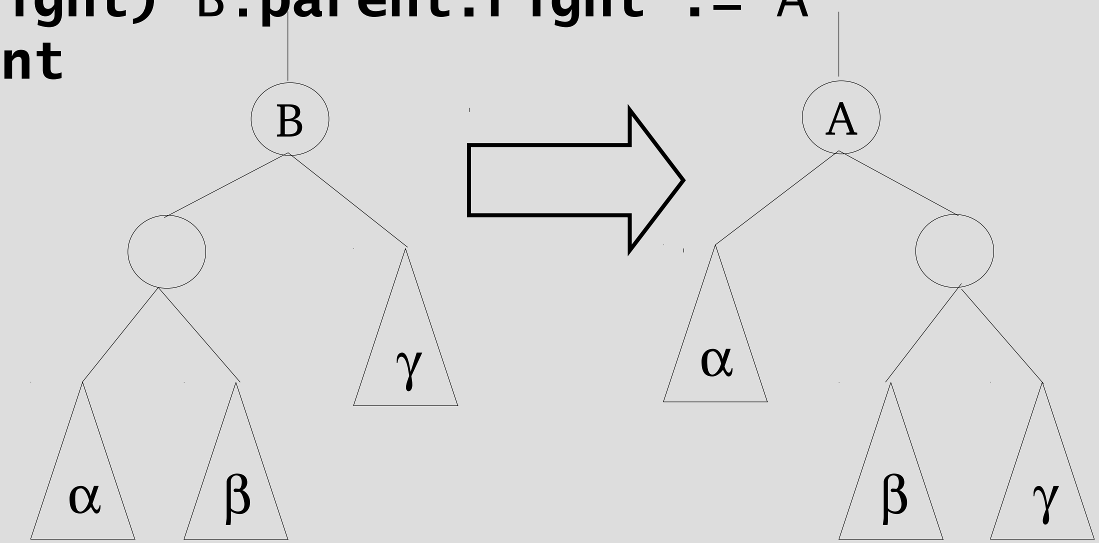
- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

Rotation



Right Rotation

```
RightRotate(B)
A := B.left
B.left := A.right
B.left.parent := B
if (B = B.parent.left) B.parent.left := A
if (B = B.parent.right) B.parent.right := A
A.parent := B.parent
A.right := B
B.parent := A
```



The Effect of a Rotation

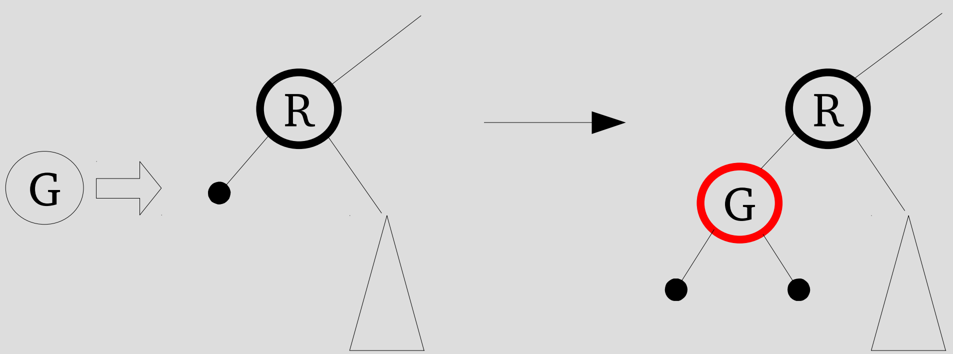
- Maintains inorder key ordering
  - $\forall a \in \alpha, b \in \beta, c \in \gamma$
    we can state the invariant
    - $a <= A <= b <= B <= c$
- After right rotation
  - Depth(α) decreases by 1
  - Depth(β) stays the same
  - Depth(γ) increases by 1
- Left rotation: symmetric
- Rotation takes *O(1)* time

Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

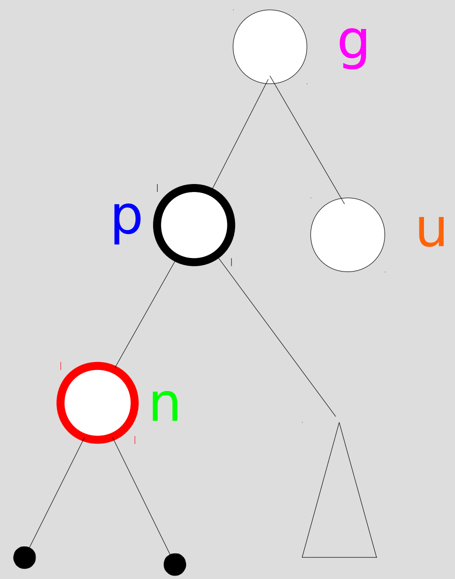
Insertion in the RB-Trees

```
RBInsert(T,n)
Insert n into T using the binary search tree
insertion procedure
n.left := NULL
n.right := NULL
n.color := red
n.parent = R;
RBInsertFixup(n)
```



Insertion

- Let
  - n = the new node
  - p = **n.parent**
  - g = **p.parent**
- In the following assume:
  - p = **g.left**
  - u = **g.right (uncle)**
- **Case 0:** p**.color** = black
  - No properties of the tree violated → done.



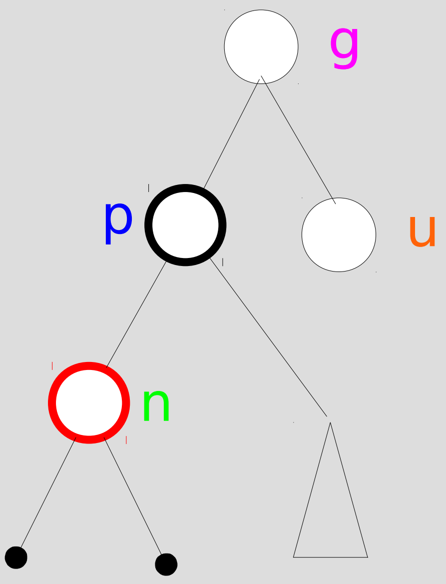
Insertion: remark

- Let
  - n = the new node
  - p = n.**parent**
  - g = p.**parent**
- If there is no parent p

  → n is the new root

  → n.color = black
- If there is no grandparent g

  → p is the root → p is black → case 0



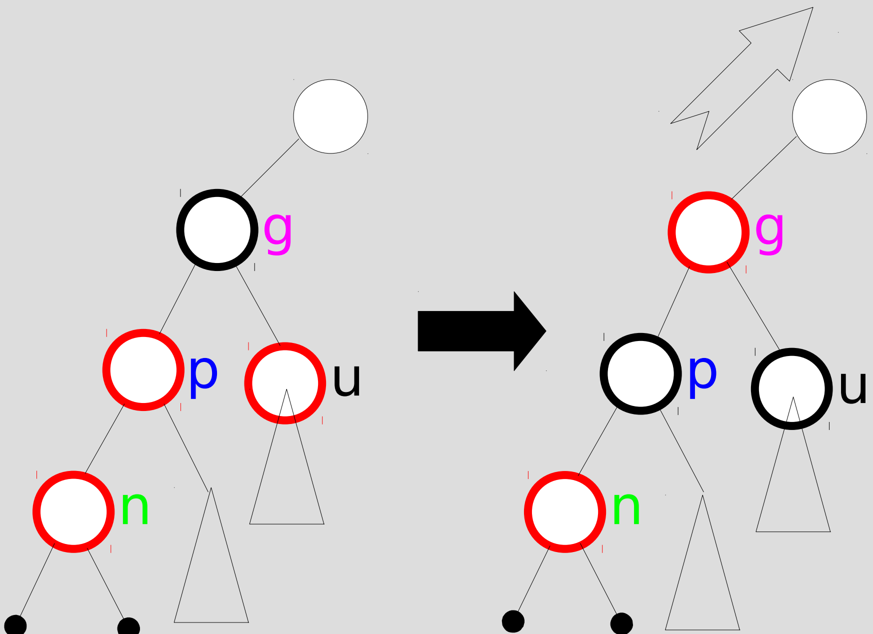
Insertion

- Hereafter we assume p is a left child p=g.left (swap right w. left otherwise)
- Three cases (p.color is red):

0. p.color = black → no violation → do nothing
1. n's uncle u is red
2. n's uncle u is black and n is a right child
3. n's uncle u is black an n is a left child

Insertion: Case 1

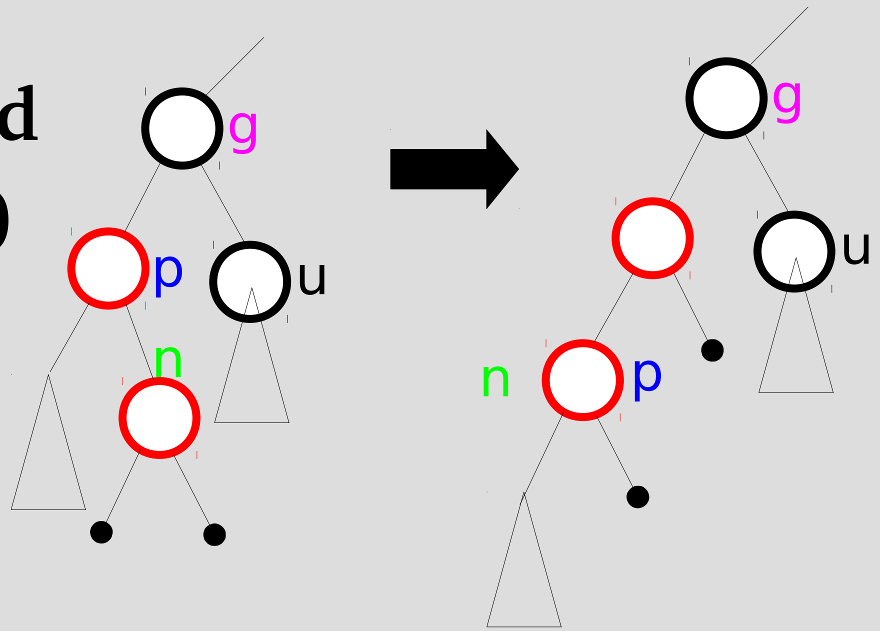
- Case 1
  - **n's uncle u is red**
- Action
  - `p.color := black`
  - `u.color := black`
  - `g.color := red`
  - `n := g`



- the tree rooted at g is balanced (black depth of all descendants unchanged)
- If root is red, make it black → no violation

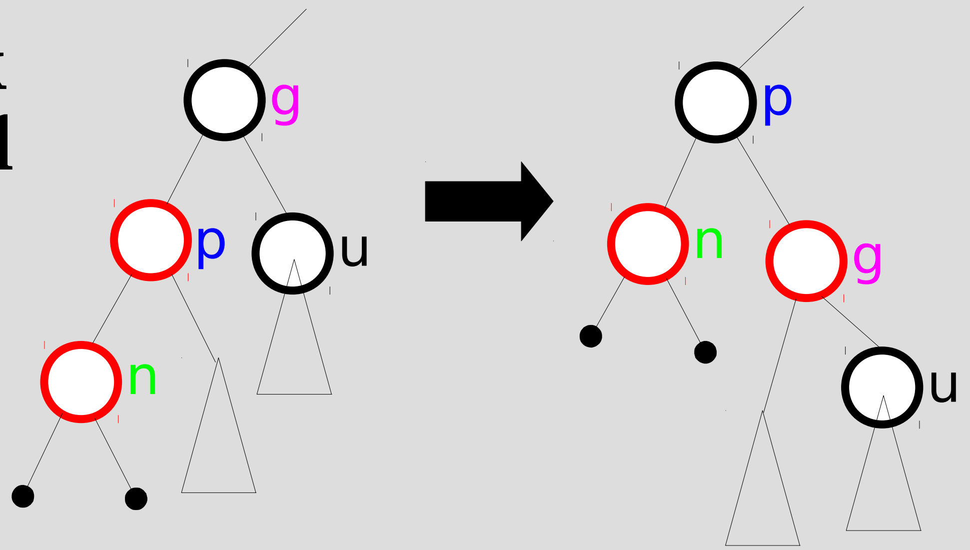
Insertion: Case 2

- Case 2
  - **n's uncle u is black and n is a right child**
  - **•(g,p,n not in a line)**
- Action
  - `LeftRotate(p)`
  - `n := p`
- Note
  - The result is a case 3.



Insertion: Case 3

- Case 3
  - **n's uncle u is black and n is a left child**
    - **(g,p,n in a line)**
- Action
  - `p.color := black`
  - `g.color := red`
  - `RightRotate(g)`
- Note: the tree rooted at g is balanced (black depth of all descendents unchanged).



Insertion: Mirror cases

- All three cases are handled analogously if p is a right child.
- Exchange *left* and *right* in all three cases.

Insertion: Case 2 and 3 mirrored

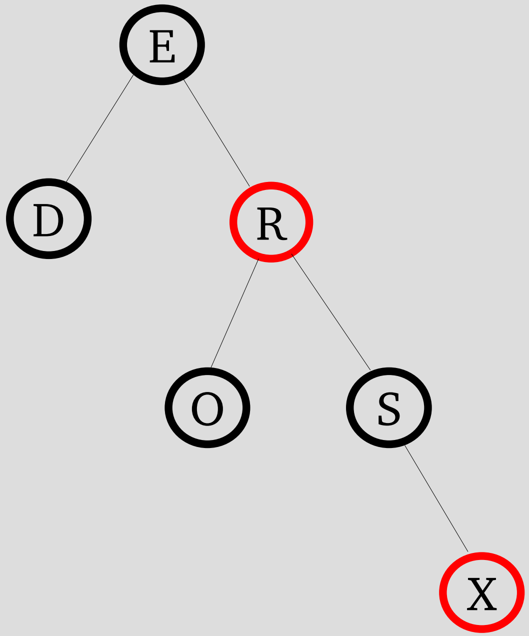
- Case 2m
  - **n's uncle u is black and n is a *left* child**
  - Action
    - `*Right*Rotate(p)`
    - `n := p`
- Case 3m
  - **n's uncle u is black and n is a *right* child**
  - Action
    - `p.color := black`
    - `g.color := red`
    - `*Left*Rotate(g)`

Insertion Summary

- If two red nodes are adjacent, we do either
  - **a restructuring** (with one or two rotations) and **stop** (cases 2 and 3), or
  - recursively **propagate** red upwards (case 1)
  - if finally the root is red, make it black → no violation
- A **restructuring** takes constant time and is performed at most once. It reorganizes an off-balanced section of the tree
- **Propagations** may continue up the tree and are executed *O(log n)* times (height of the tree)
- The running time of an insertion is ***O(log n)***.

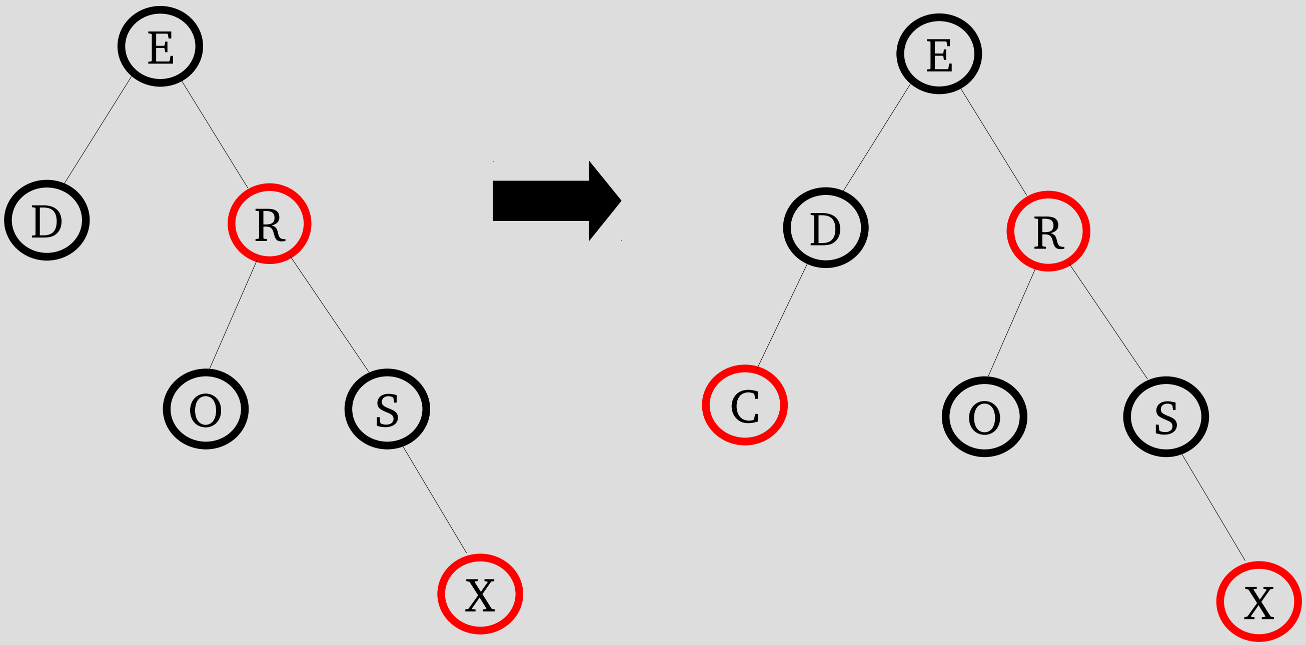
An Insertion Example

- Inserting "REDSOX" into an empty tree

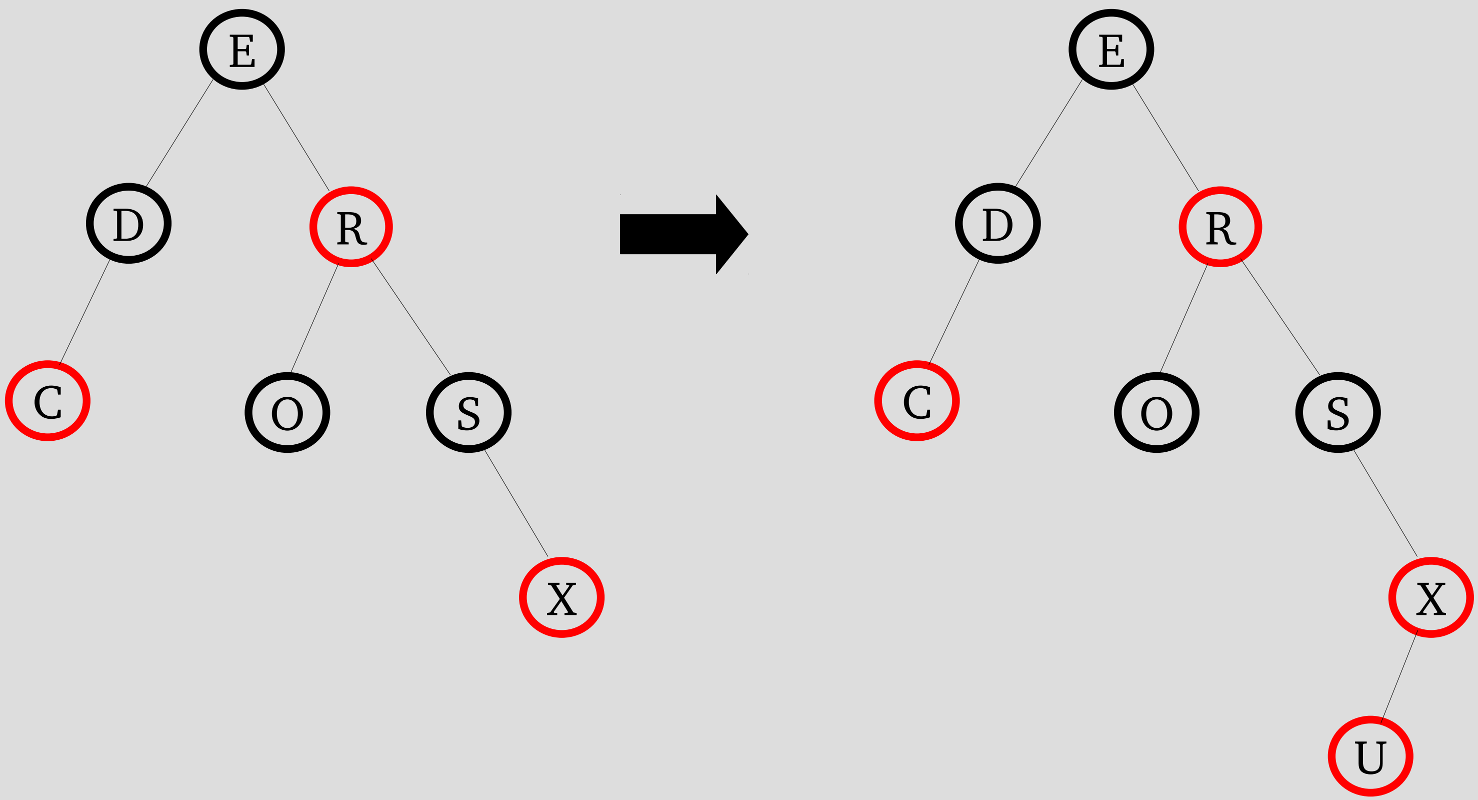


- Now, let us insert "CUBS"

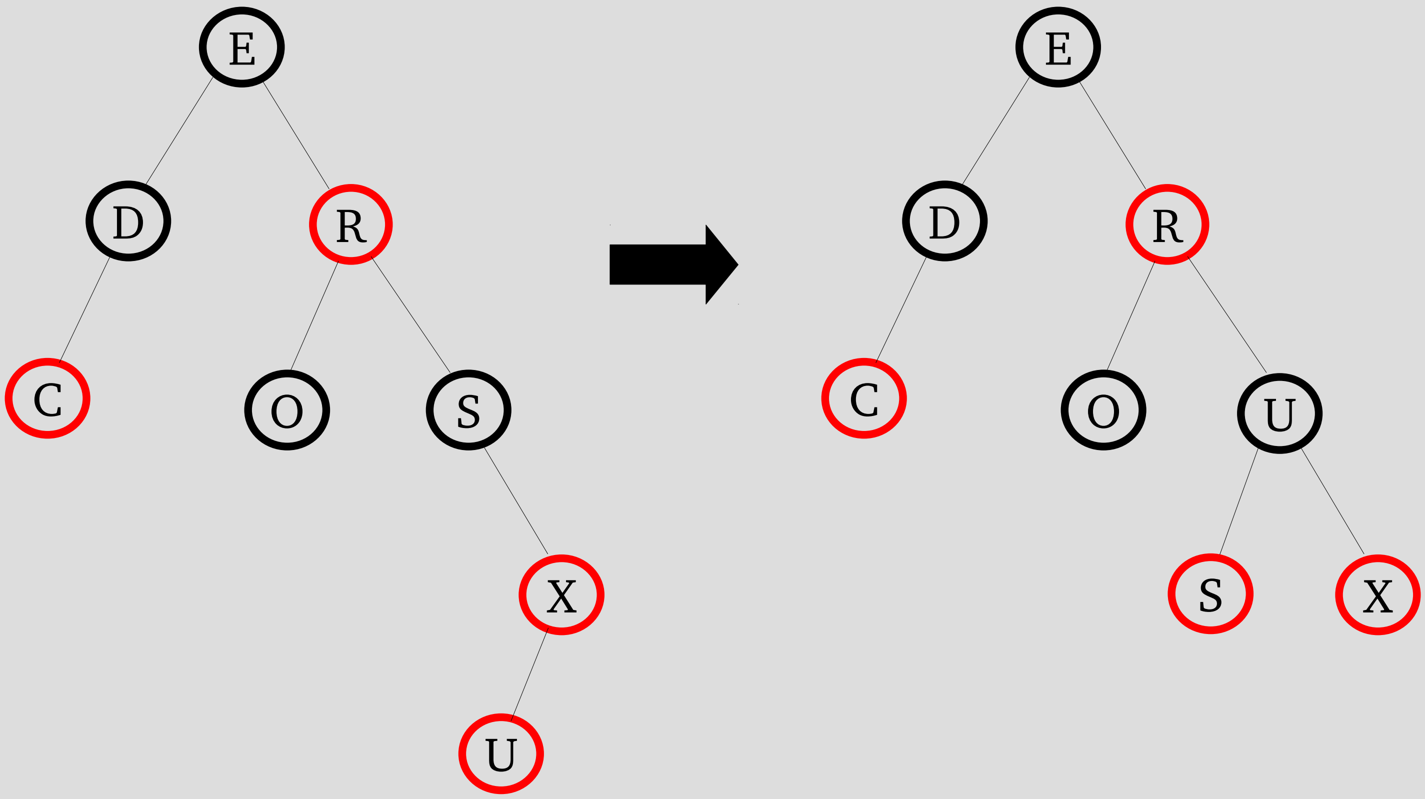
Insert C (case 0)



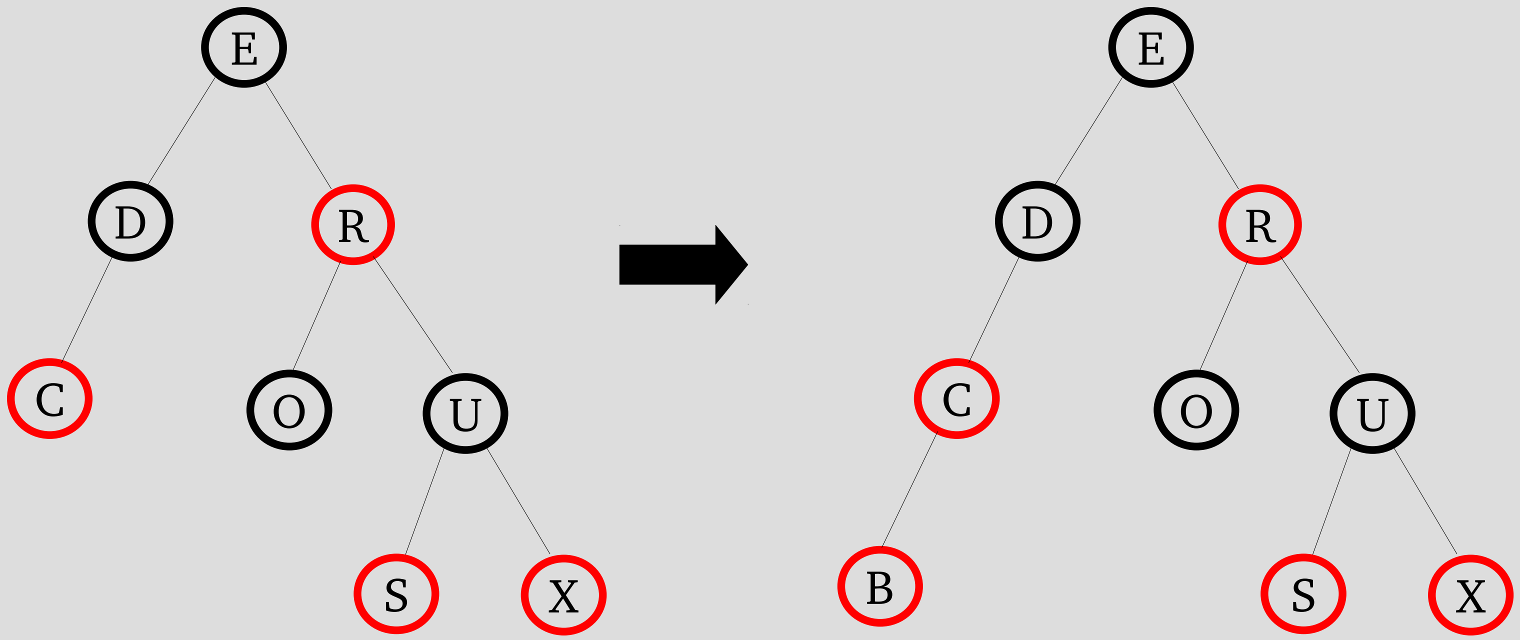
Insert U (case 2, mirror)



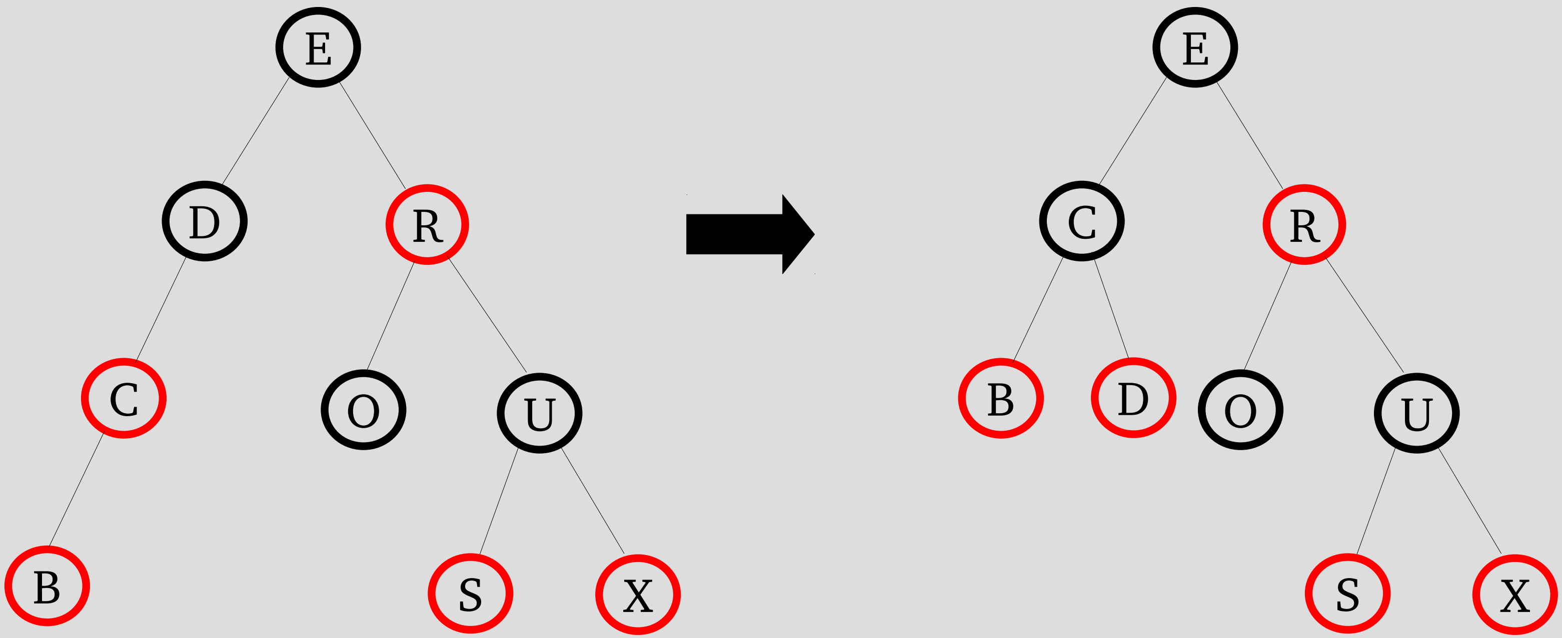
Insert U/2



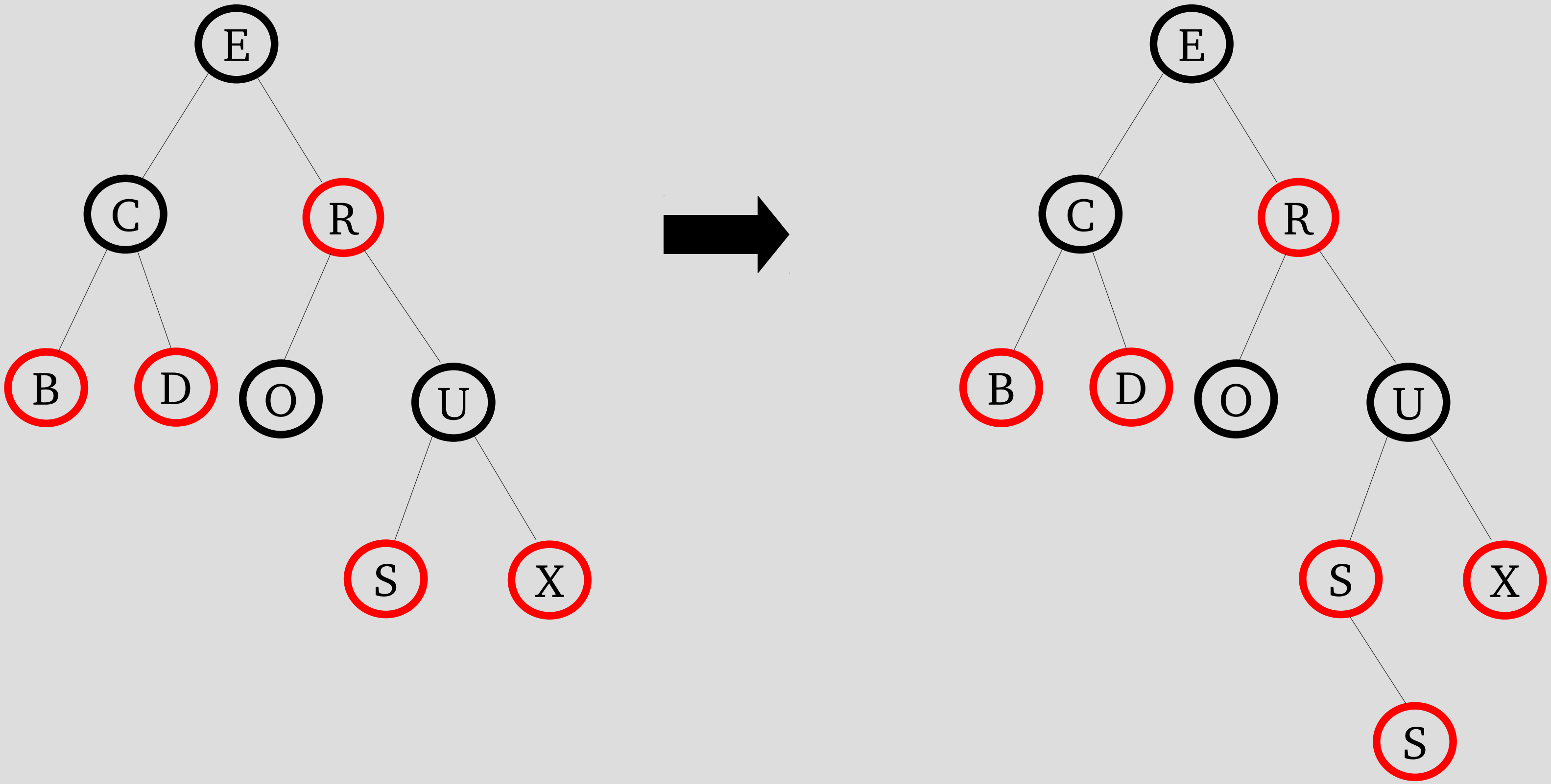
Insert B (case 3)



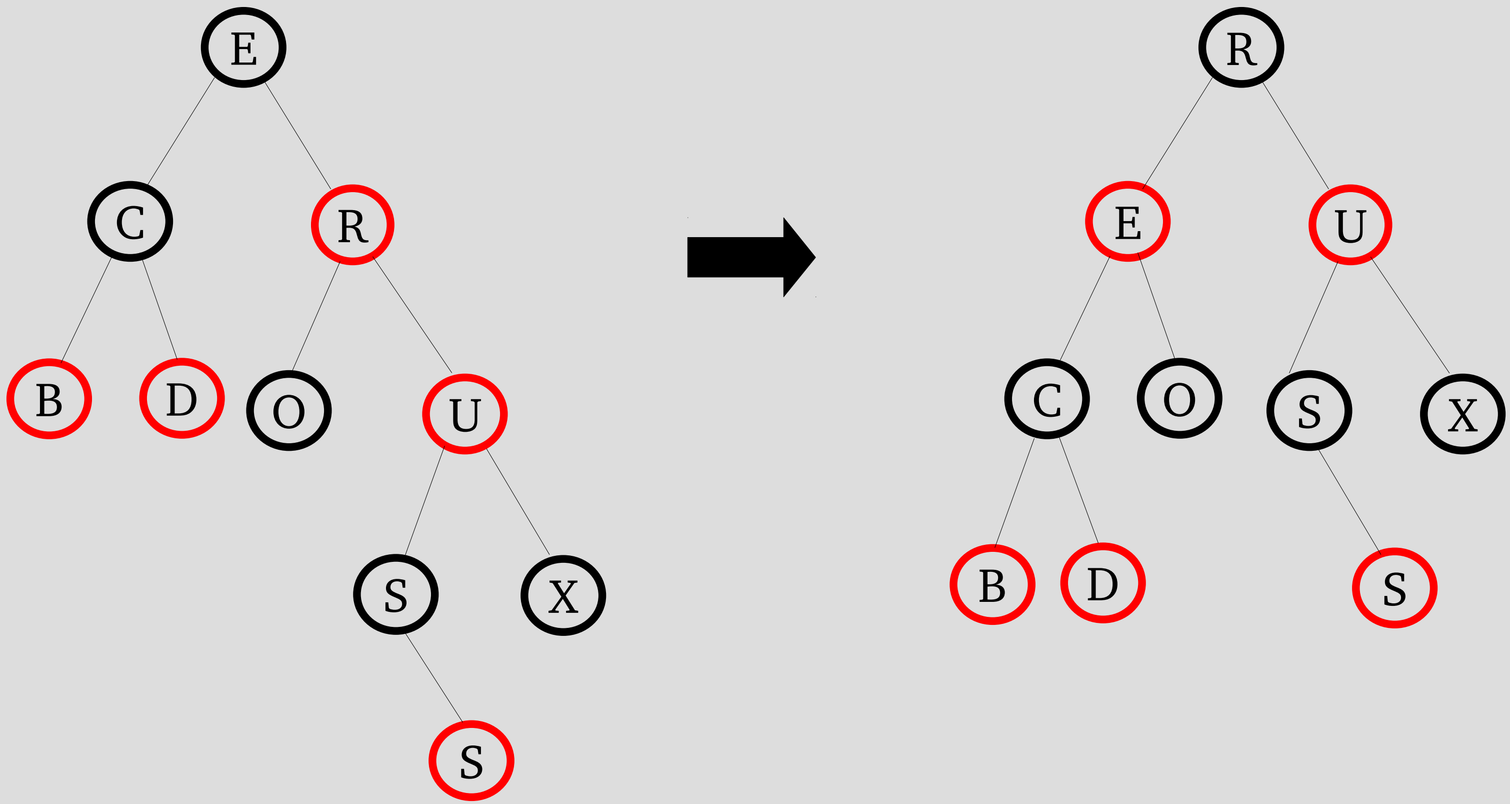
Insert B/2



Insert S (case 1)



Insert S/2 (case 3 mirror)

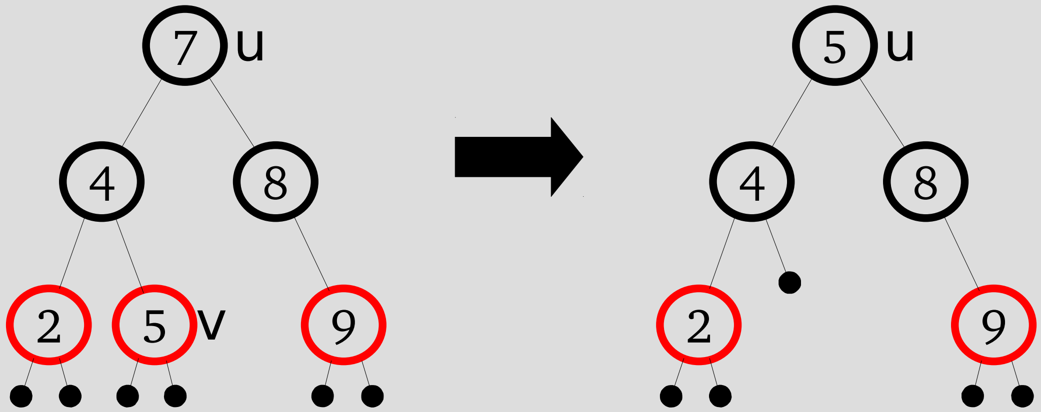


Data Structures and Algorithms
Week 6

- Binary Search Trees
  - Tree traversals
  - Searching
  - Insertion
  - Deletion
- Red-Black Trees
  - Properties
  - Rotations
  - Insertion
  - Deletion

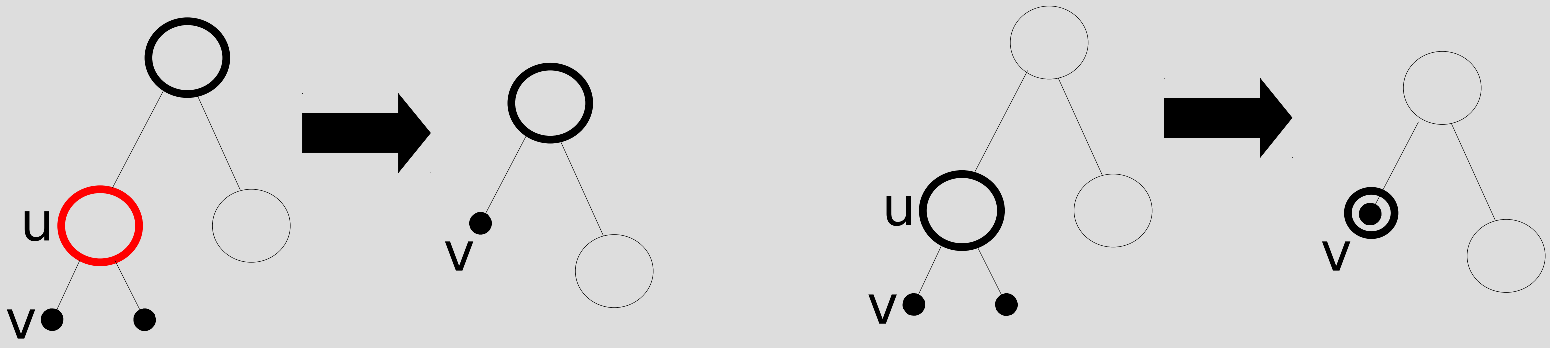
Deletion

- We first apply binary search tree deletion.
  - We can easily delete a node that has at least one *nil* child
  - If the key to be deleted is stored at a node u with two children, we replace its content with the content of the largest node v of the left subtree and delete v instead.



Deletion Algorithm

1. Remove u
2. If u.**color** = red, we are done. Else, assume that v (replacement of u) gets *additional black color:*
   - If v.**color** = red then v.**color** := black and we are done!
   - Else v' s color is **“double black”**.



Deletion Algorithm/2

- How to eliminate double black edges?
  - The intuitive idea is to perform a **color compensation**
    - Find a red edge nearby, and change the pair (**red**, **double black**) into (**black, black**)
  - Two cases: **restructuring** and **recoloring**
  - Restructuring resolves the problem locally, while recoloring may propagate it upward.
- Hereafter we assume v is a left child (swap right and left otherwise)

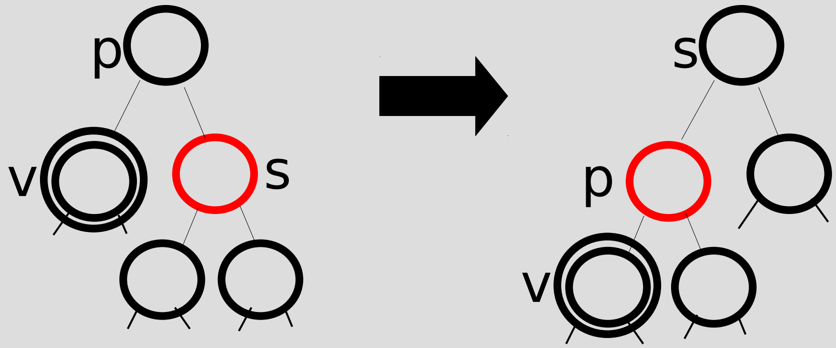
Eliminating double-back nodes

- Hereafter we assume v is a left child (swap right with left otherwise)
- Four cases:

0. v is the root → do nothing
1. v's sibling s is red
2. [recoloring] v's sibling s is black and both children of s are black
3. v's sibling s is black, s's left child is red, and s's right child is black
4. [restructuring] v's sibling s is black, and s's right child is red (regardless s's left child)

Deletion: Case 1

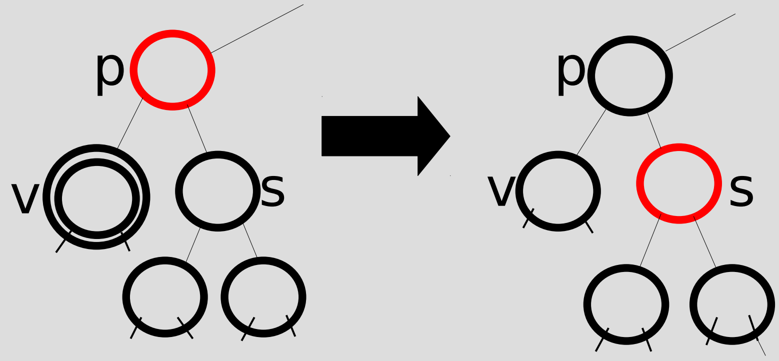
- Case 1
  - **v's sibling s is red (→ p is black)**
- Action
  - `s.color = black`
  - `p.color = red`
  - *Left*`Rotation(p)`
  - `s = p.right`



- Note
  - This is now a case 2, 3, 4

Deletion Case 2 (recoloring)

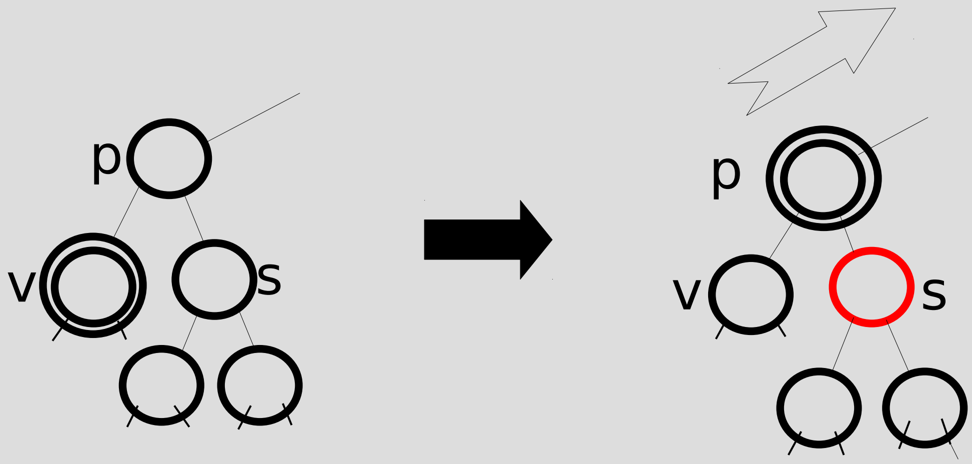
- Case 2
  - **v's sibling s is black and both children of s are black**
- Action
  - `s.color := red`
  - `v = p`



- Note
  - We reduce the black depth of both subtrees of p by 1. P becomes "more" black.

Deletion: Case 2 (recoloring)

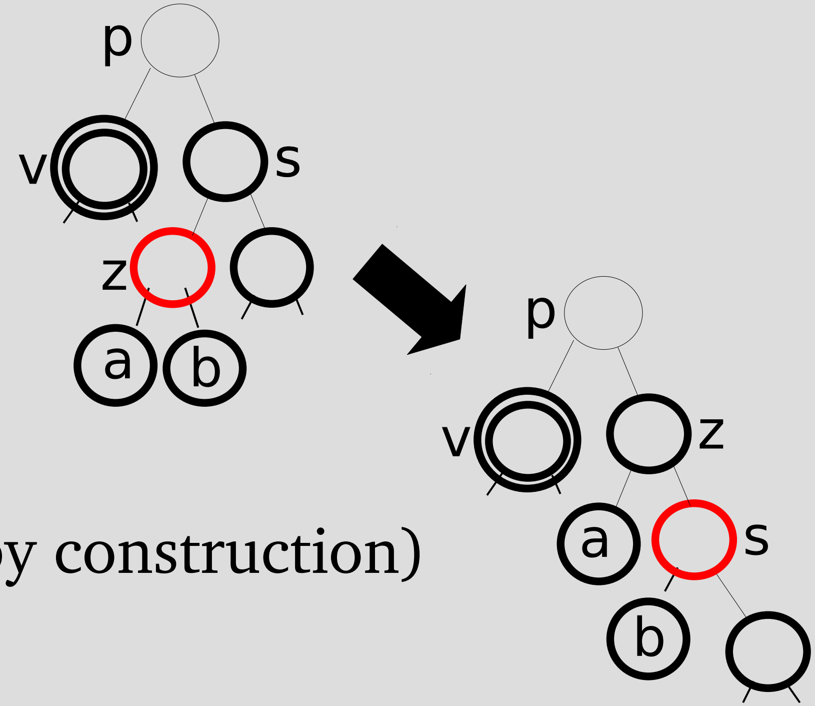
- If parent becomes **double black**, continue upward.



Deletion: Case 3

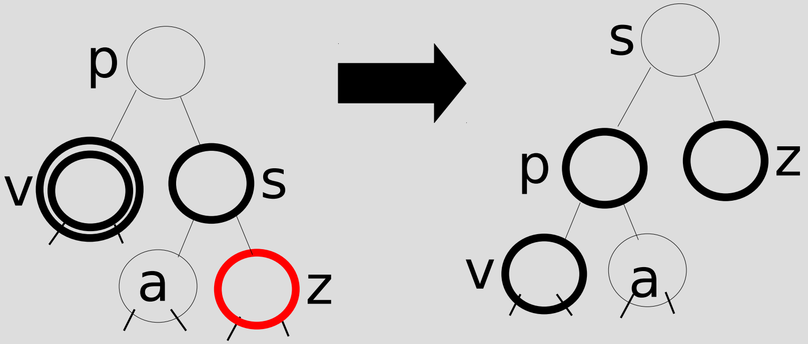
- Case 3
  - **v's sibling s is black, s's left child is red, and s's right child is black.**
- Action
  - `s.left.color = black`
  - `s.color = red`
  - `RightRotation(s)`
  - `s = p.right`
- Note: This is now a case 4

  (z's children must be black by construction)



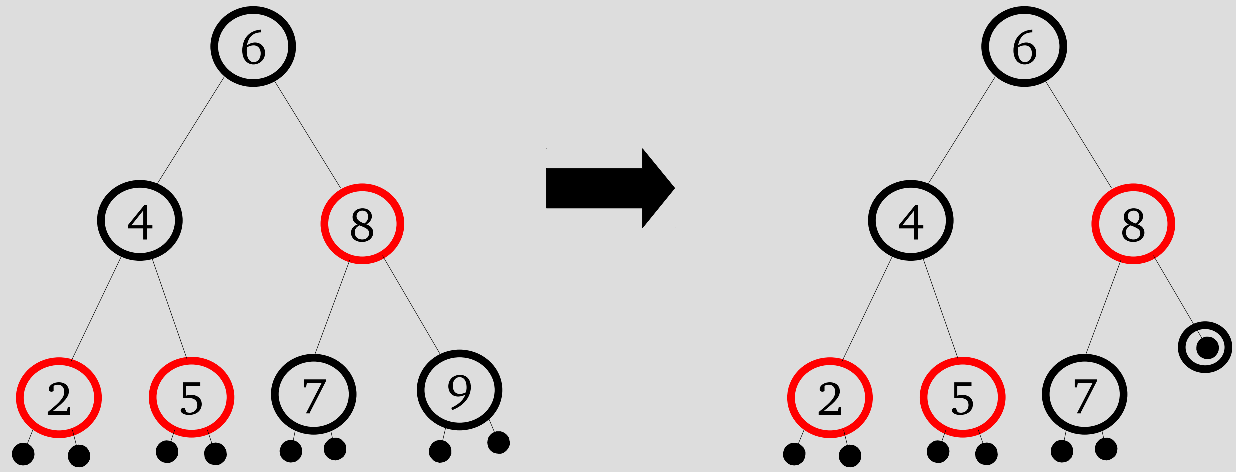
Deletion: Case 4 (restructuring)

- Case 4
  - **v's sibling s is black and s's right child is red (regardless s's left child)**
- Action
  - `s.color = p.color`
  - `p.color = black`
  - `s.right.color = black`
  - `LeftRotate(p)`
  - `v.color = black // single`



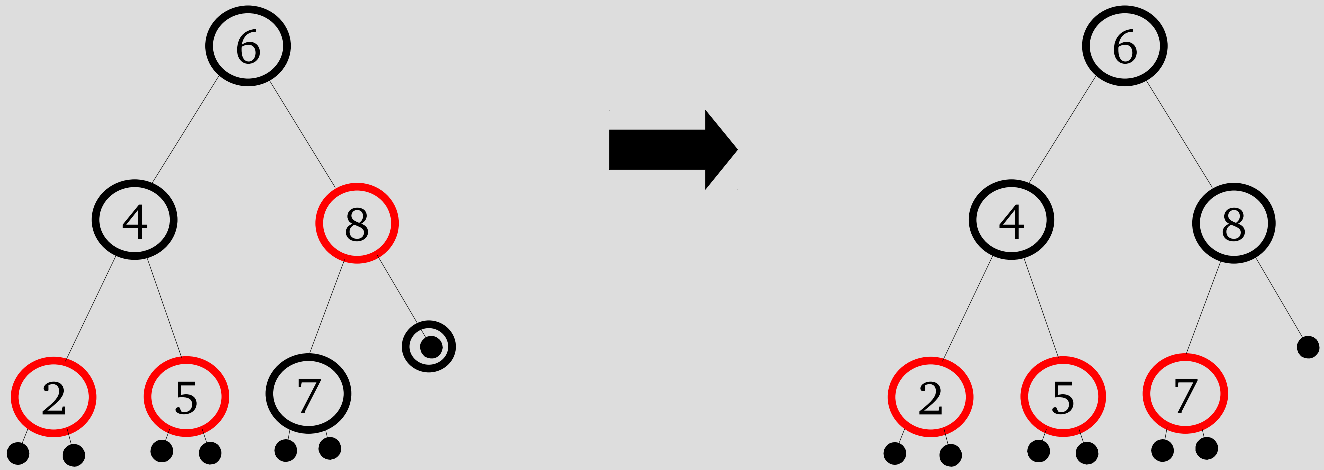
- Note
  - Terminates after restructuring

Delete 9

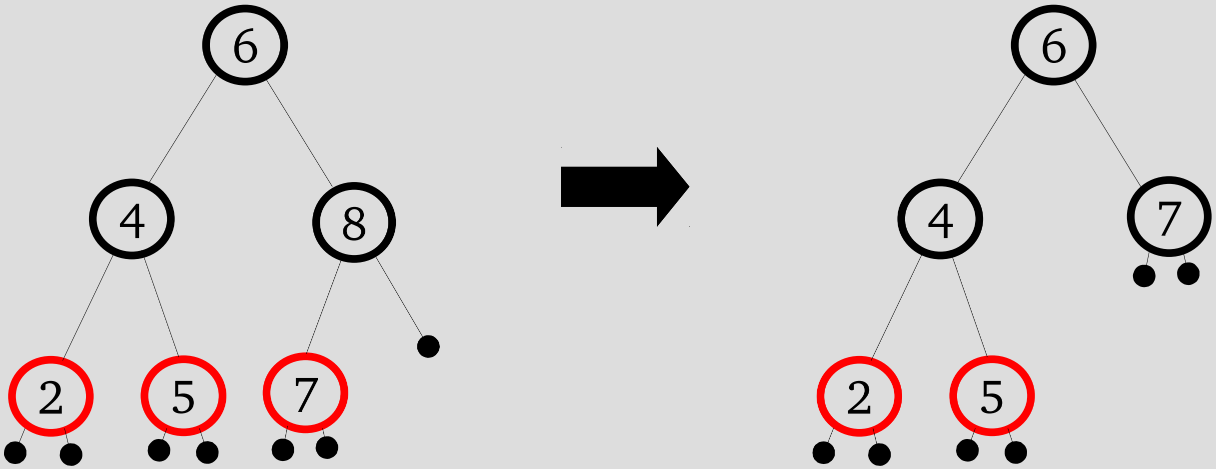


Delete 9/2

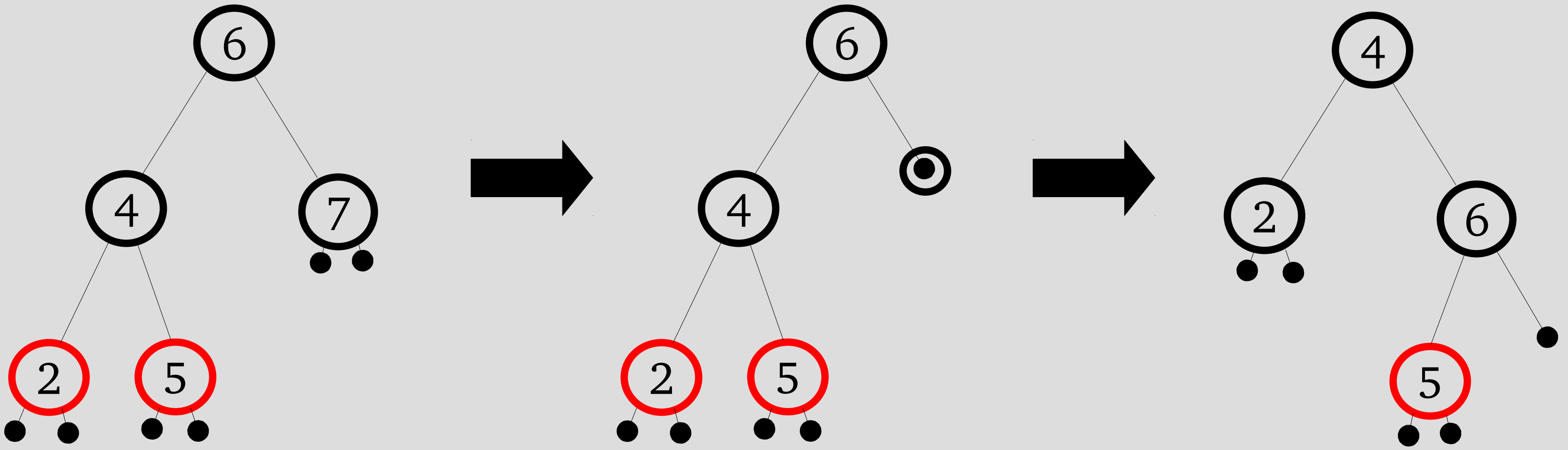
- Case 2 (sibling is black with black children) – recoloring



Delete 8



Delete 7: restructuring



How long does it take?

- Deletion in a RB-tree takes $O(\log n)$
  - Maximum three rotations and $O(\log n)$ recolorings

Suggested exercises

- Add left-rotate and right-rotate to the implementation of binary trees
- Implement a red-black search tree with the following functionalities:
  - (...), insert, delete

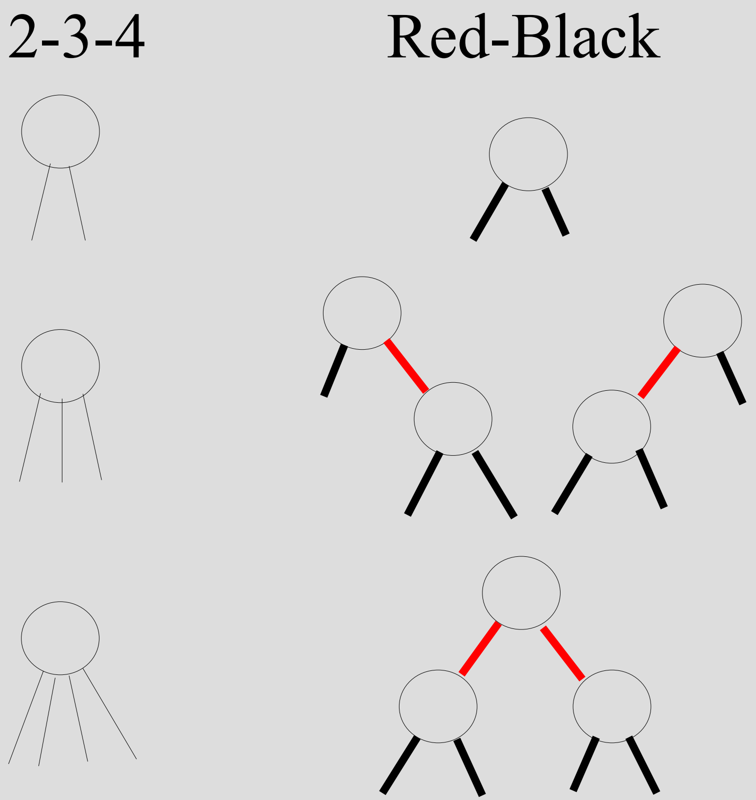
Suggested exercises/2

Using paper & pencil:

- draw the RB-trees after each of the following operations, starting from an empty tree:
  1. Insert 1,2,3,4,5,6,7,8,9,10,11,12
  2. Delete 12,11,10,9,8,7,6,5,4,3,2,1
- Try insertions and deletions at random

Other Balanced Trees

- Red-Black trees are related to 2-3-4 trees (non-binary)
- AVL-trees have simpler algo-rithms, but may perform a lot of rotations



Next Week

- Hashing