



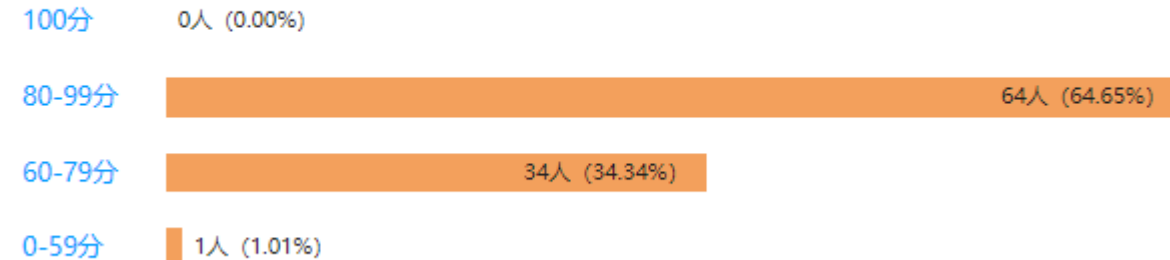
SEARCH TREES

School of Artificial Intelligence

ON EXAMS

- Mid-term
 - One page of key knowledge points
 - Difficulty: moderate
 - Average completion time: 60 minutes
- Final exam
 - Close book
 - Diverse types of questions
 - 50% of your final grade

分数分布



未参加考试及待批阅的学生均以0分计算

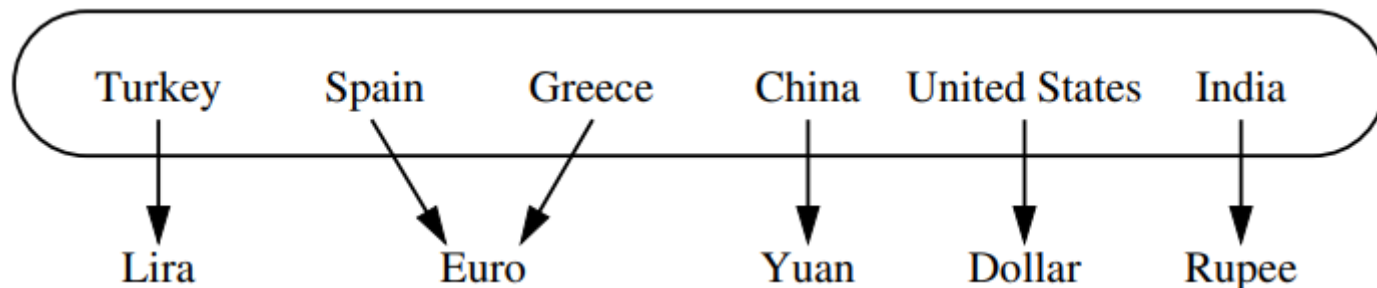


PREVIOUSLY ON DS&A

- Maps and Dictionaries
- Hash tables
- Hash functions
 - Hashing
 - Compressing
- Collision handling
 - Linear probing
 - Quadratic probing
 - Double hashing
- Skip Lists

MAPS AND DICTIONARIES

- Key->Value pairing
 - Unique association
- Most significant data structure in any programming language
- Often known as **associative arrays** or **maps**
- Keys are (assumed) to be unique, values are not necessarily unique
- Python: dict class
- Use key as 'index'
- Indices need not be consecutive nor numeric



HASH TABLES (哈希表)

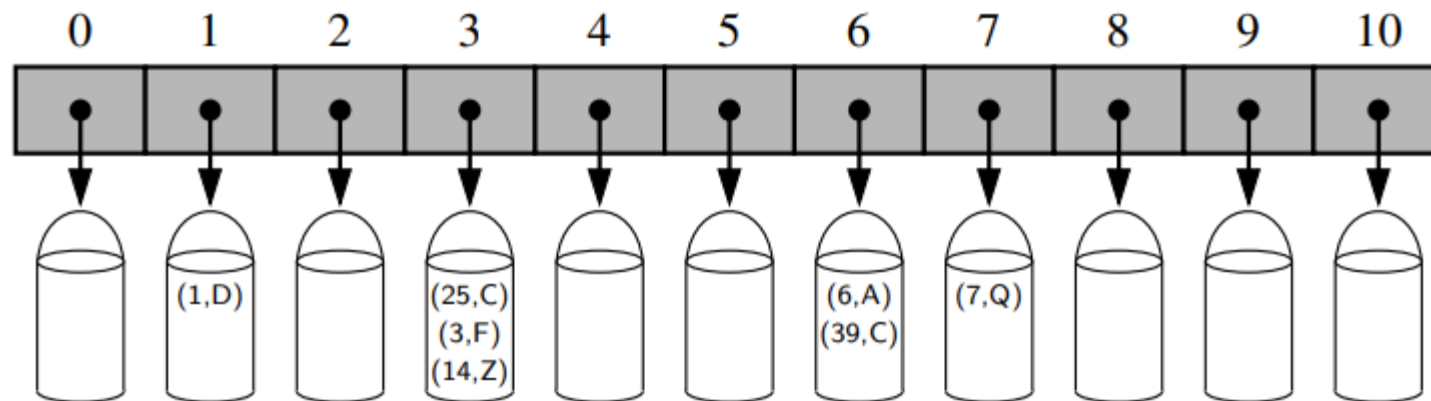
- Most practical data structures for implementing a map
- A map M supports the abstraction of using keys as indices with a syntax such as $M[k]$.
- Assume a map with n items uses integer keys from 0 to $N-1$ for some $N \geq n$
- We can represent the map like this:

0	1	2	3	4	5	6	7	8	9	10
	D		Z			C	Q			

- `__getitem__`, `__setitem__` and `__delitem__` become $O(1)$
- Problems?

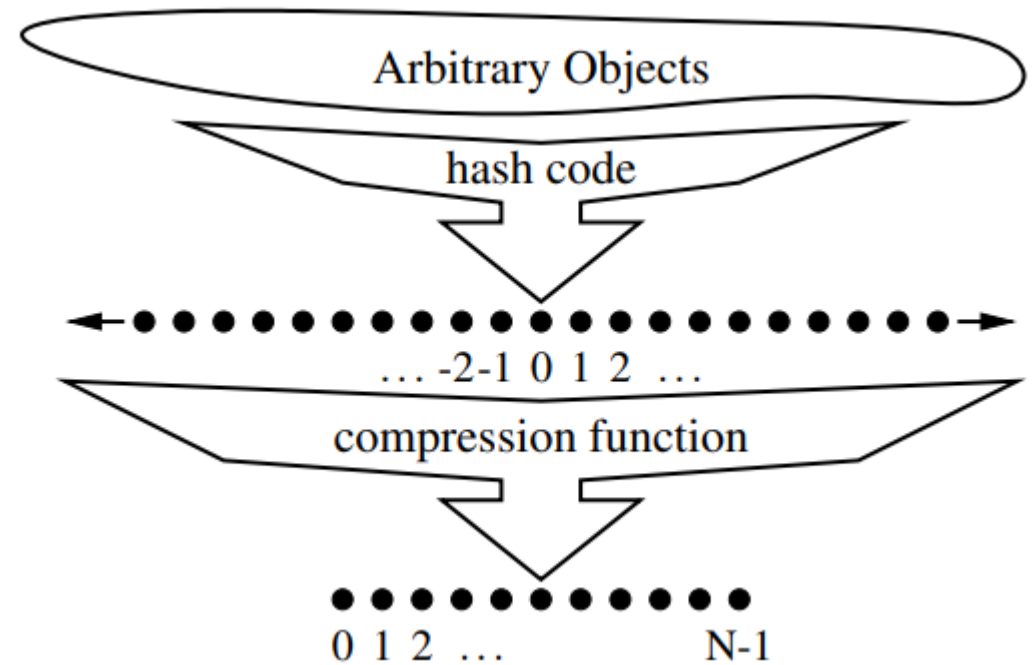
HASH TABLES (哈希表)

- Problems
 - Keys n may not be continuous, therefore an array for the map may have size $N \gg n$
 - A map's key can be other data types, not just integers
- Solution: hash function to map keys to corresponding indices in a table
- Ideally, keys will be distributed in the range from 0 to $N-1$
- But, there may be two or more distinct keys that get mapped to the same index
- Bucket array



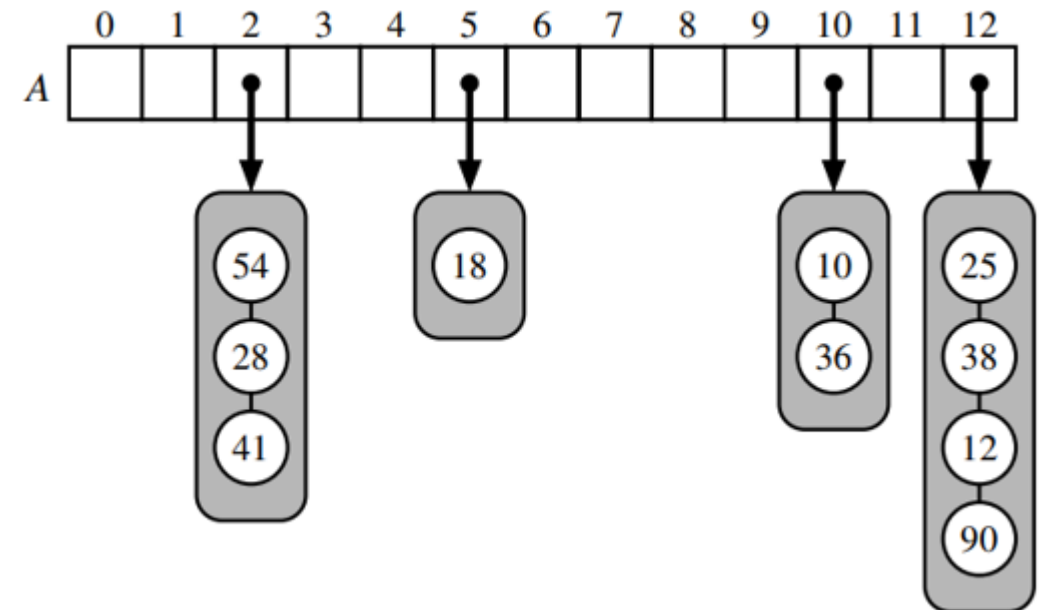
HASH FUNCTION (哈希函数)

- Hash function: $h(k)$
- **Hashing**: produce a *hash code* that maps a key k to an integer
- **Compressing**: maps the hash code to an integer within a range of indices $[0, N-1]$
- Why the separation?
 - Independence: hashing is independent of a specific hash table size
 - OO design: hash functions can be overridden



COLLISION HANDLING

- Separate chaining
- Operations on an individual index (bucket) take time proportional to the size of the bucket
- “Good” hash function:
 - expected size of a bucket: n/N
 - n = number of items in the map
 - N = capacity of the bucket array
 - Core map operations run in $O(\text{ceiling}(n/N))$
- Load factor (负载因子) : $\lambda = n/N$
- When λ is $O(1)$, operations on the hash table run in $O(1)$

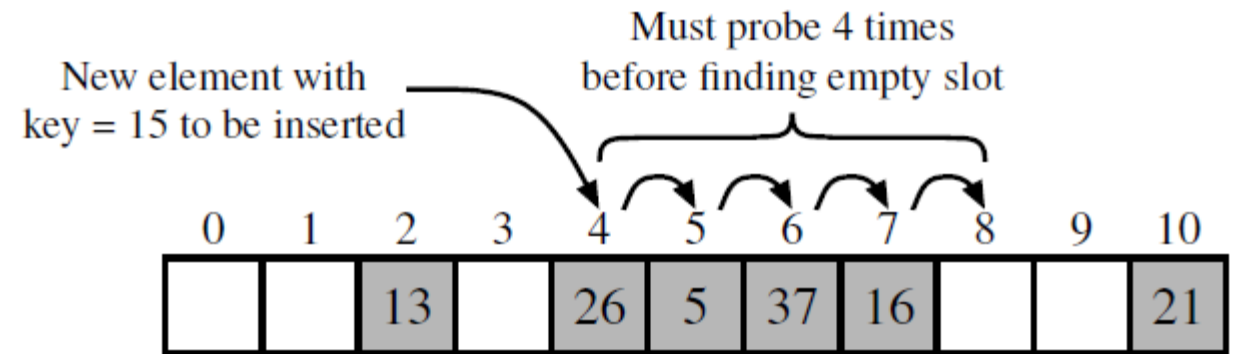


COLLISION HANDLING

- Separate chaining （分离链表）：
 - Advantage: simple implementation
 - Disadvantage: relies on auxiliary data structure - list to hold items with colliding keys
- Alternative approach: open addressing （开放寻址）
 - Always store each item directly in a table slot
 - No auxiliary structures are used
 - Load factor is ways at most 1 and items are stored directly in the cells of the bucket array

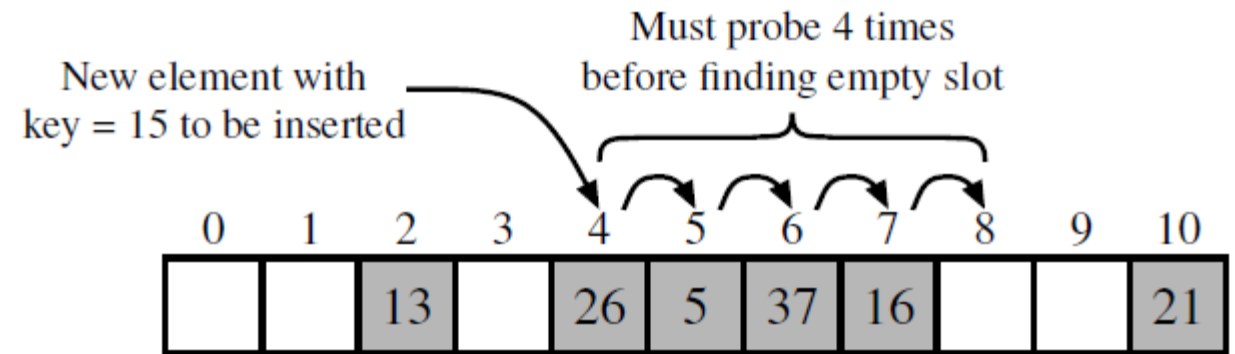
LINEAR PROBING (线性探索)

- Linear probing
- Insert an item (k, v) into a bucket $A[j]$
- If $a[j]$ is occupied, $j = h(k)$, then try $A[(j+1) \bmod N]$
- If $A[(j+1) \bmod N]$ is occupied, try $A[(j+2) \bmod N]$, so on
- Need to change the implementation of functions such as `__getitem__`, `__setitem__`, `__delitem__`



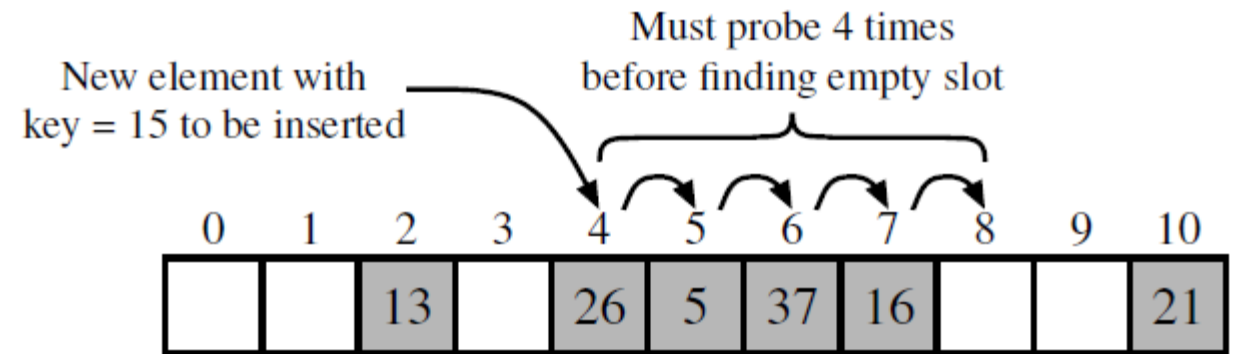
QUADRATIC PROBING (二次探索)

- Iteratively tries $A[h(k) + f(i) \bmod N]$ for $i = 0, 1, 2, \dots$, $f(i) = i^2$
- Spreads the probing distance over the length N
- Deletion – same strategy as linear probing
- Problems again: secondary clustering (二次聚集)
- When the bucket array is half full, or N is not a prime, quadratic probing does not guarantee to find an empty slot



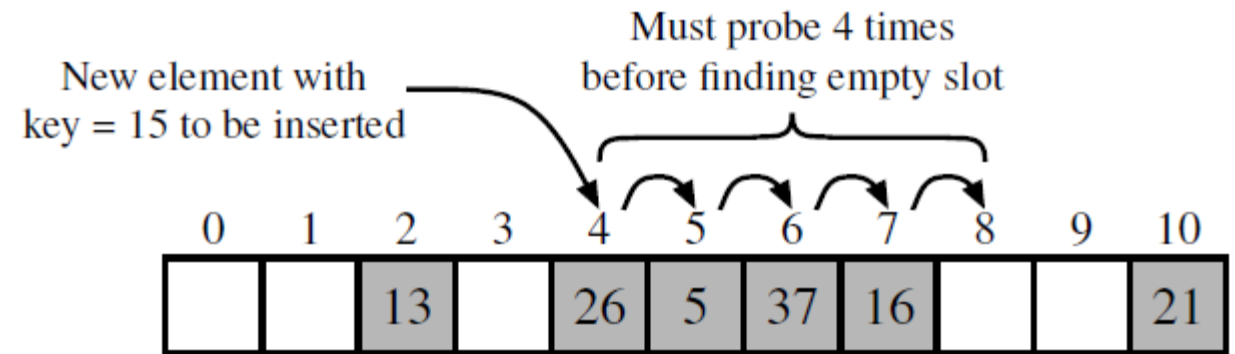
DOUBLE HASHING (二次哈希)

- Secondary hash function h'
- If h maps some key k to a bucket $A[h(k)]$ that is occupied, try $A[h(k) + f(i) \bmod N]$, for $i = 1, 2, 3, \dots$
- $f(i) = i * h'(k)$
- $h'(k)$ cannot be 0
- $h'(k) = q - (k \bmod q)$, q, N are prime numbers and $q < N$



ADDITIONAL OPEN ADDRESSING

- $A[h(k) + f(i) \bmod N]$ where $f(i)$ produces a pseudo-random number
- Repeatable random number

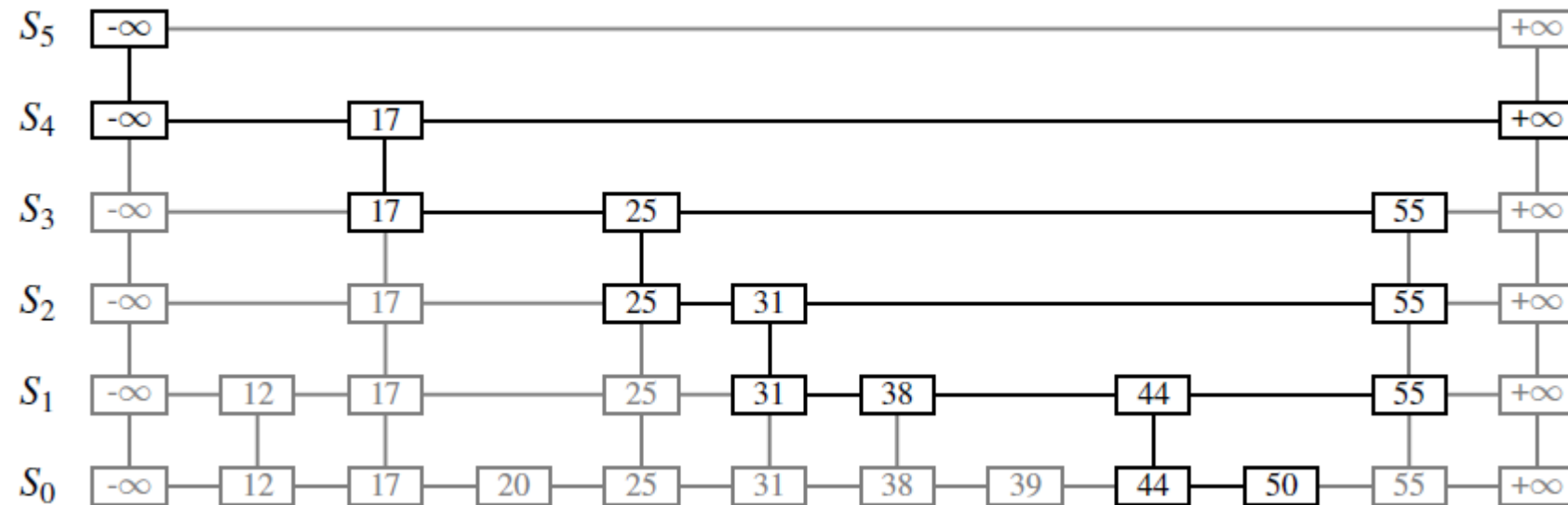


LOAD FACTORS AND REHASHING

- Load factor (负载因子) : $\lambda = n/N$ should be kept below 1
- Separate chaining: $\lambda \rightarrow 1$ the probability of a collision increases greatly
 - $\lambda < 0.9$ for separate chaining
- Open addressing
 - When λ grows beyond 0.5 and approaches 1, clusters start to show
 - Linear probing: $\lambda < 0.5$
 - Other open addressing: $\lambda < 2/3$
- What happens if an insertion causes the load factor to go beyond 0.5 for linear probing and 2/3 for other open addressing means?
- **Rehashing**: resize the table + reinsert all objects into new table
 - Resize: how?
 - new compression function -> why?

SKIP LISTS

- Search: SkipSearch(k)
 - Key k
 - Position p to point to the top left position – **start position**
1. If $S.\text{below}(p)$ is None, search is done, obtained key at position $p \leq k$. Otherwise $p = S.\text{below}(p)$
 2. Move p forward until it is at the rightmost position, such that $k(p) \leq k$
 3. Return to step 1



SKIP LISTS

Algorithm SkipSearch(k):

Input: A search key k

Output: Position p in the bottom list S_0 with the largest key such that $\text{key}(p) \leq k$

$p = \text{start}$

{begin at start position}

while $\text{below}(p) \neq \text{None}$ **do**

$p = \text{below}(p)$

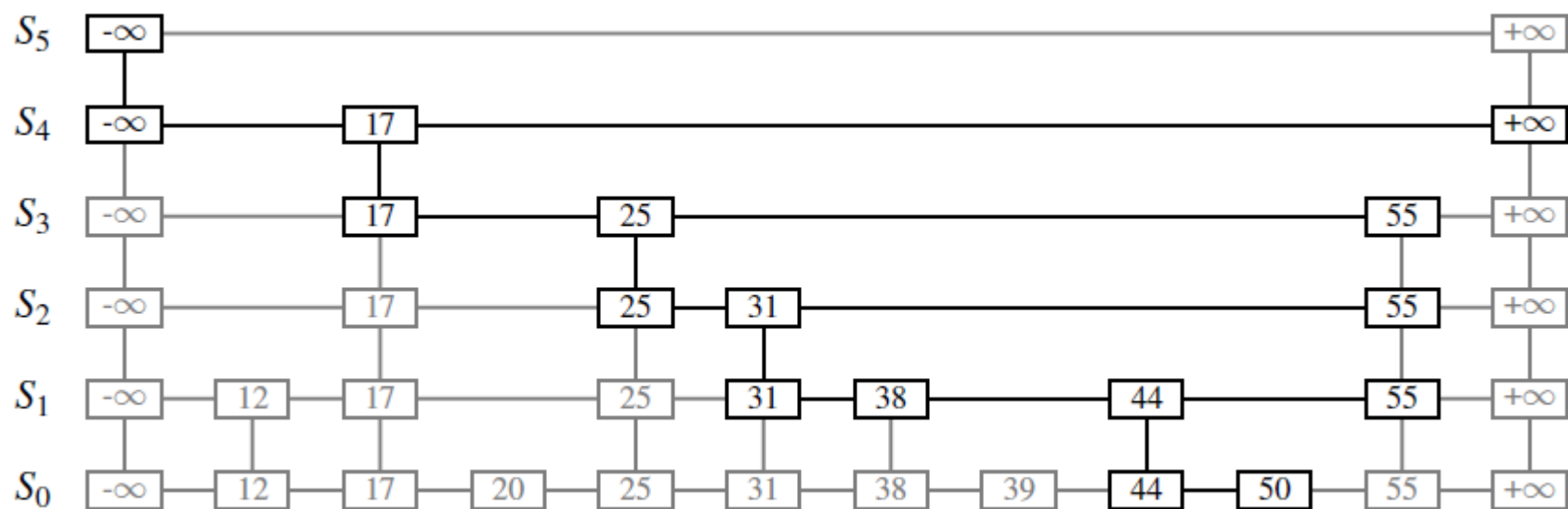
{drop down}

while $k \geq \text{key}(\text{next}(p))$ **do**

$p = \text{next}(p)$

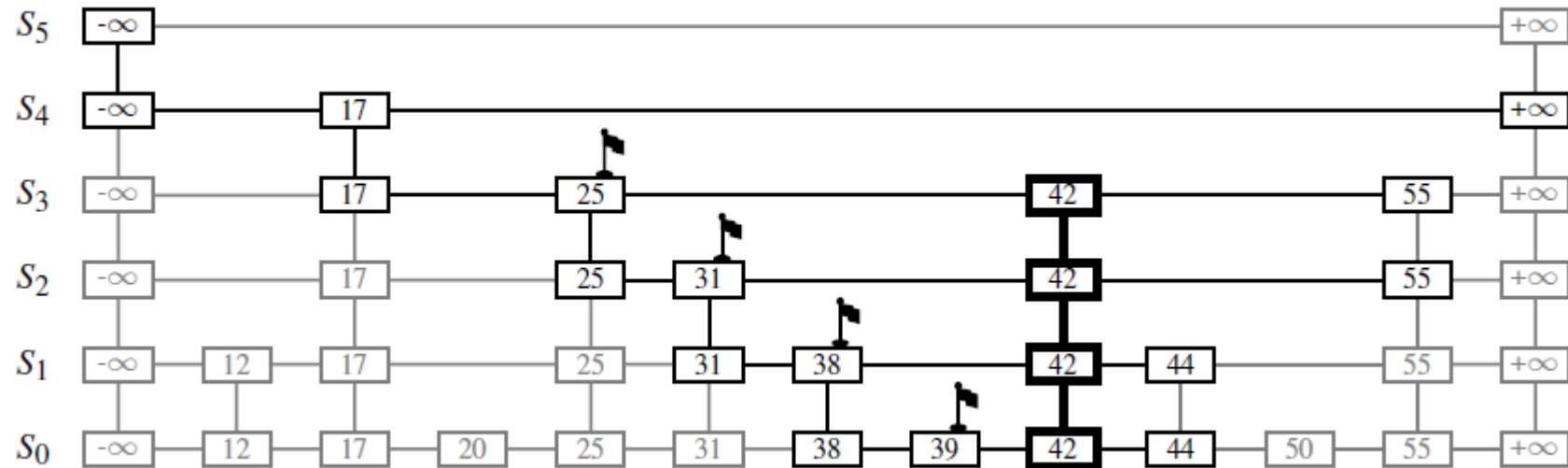
{scan forward}

return p .



SKIP LISTS

- Insertion
- Begins with $\text{SkipSearch}(k)$ to find a position p
- If $\text{key}(p) == k$, update (k, v)
- Otherwise, need to create a new **tower**
- Insert (k, v) immediately after position p within S_0
- Randomise to decide the height of the tower



SKIP LISTS

Algorithm SkipInsert(k, v):

Input: Key k and value v

Output: Topmost position of the item inserted in the skip list

$p = \text{SkipSearch}(k)$

$q = \text{None}$ { q will represent top node in n }

$i = -1$

repeat

$i = i + 1$

if $i \geq h$ **then**

$h = h + 1$ {add a new level}

$t = \text{next}(s)$

$s = \text{insertAfterAbove}(\text{None}, s, (-\infty, \text{None}))$ {grow}

$\text{insertAfterAbove}(s, t, (+\infty, \text{None}))$ {grow}

while $\text{above}(p)$ is **None** **do**

$p = \text{prev}(p)$

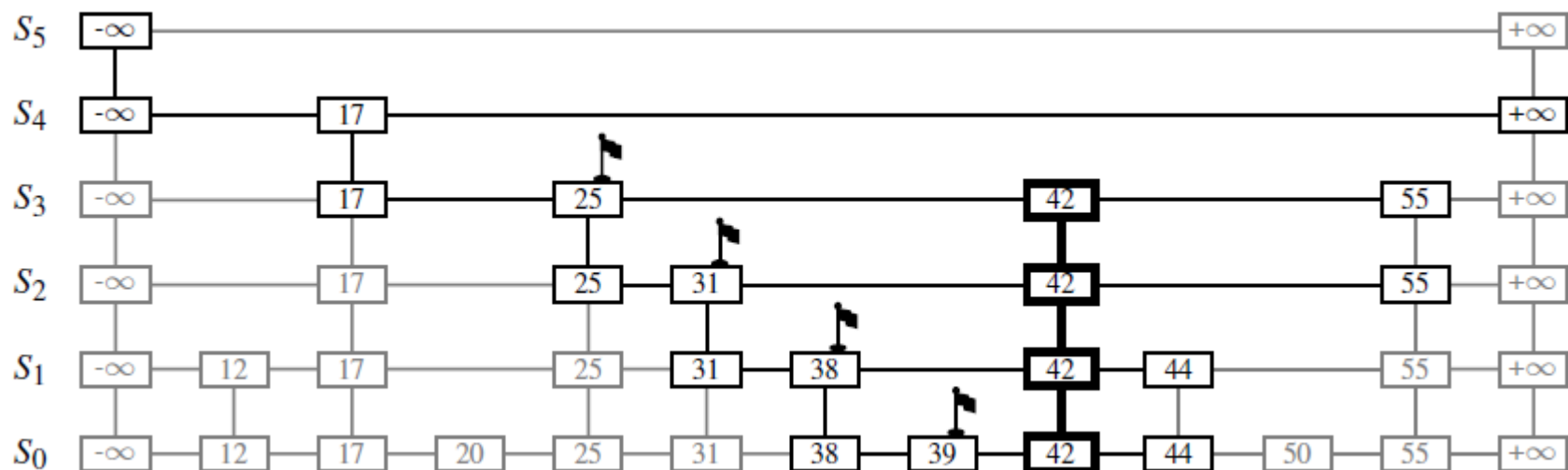
$p = \text{above}(p)$

$q = \text{insertAfterAbove}(p, q, (k, v))$

until $\text{coinFlip}() == \text{tails}$

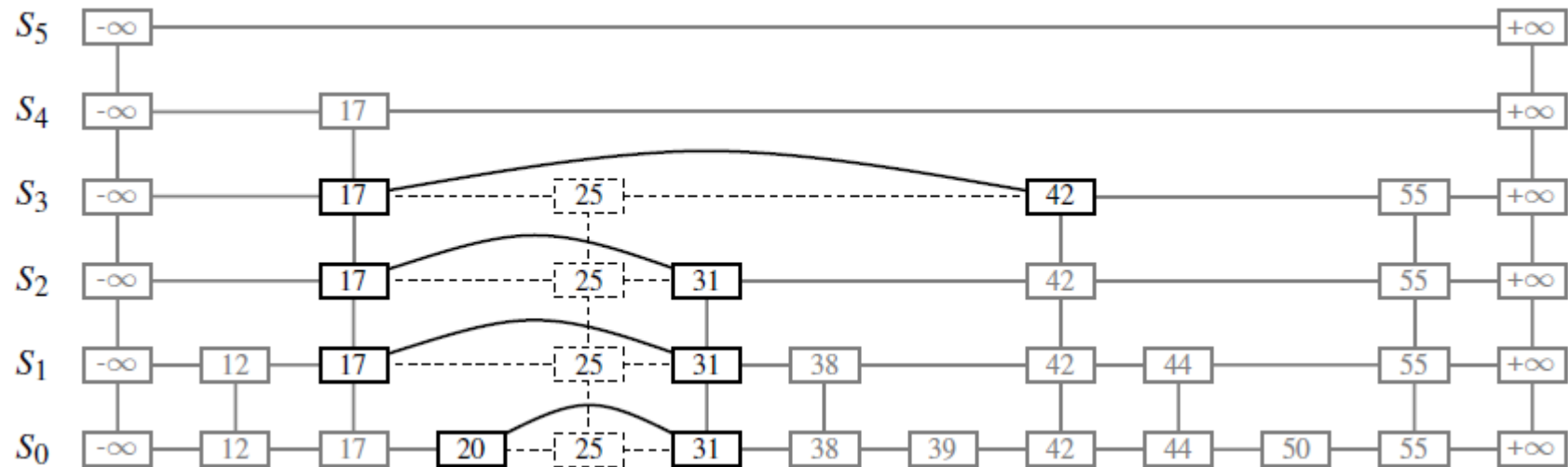
$n = n + 1$

return q



SKIP LISTS

- Removal
- Perform SkipSearch(k) to obtain p
- If p is illegal, raise error
- If p is legal, remove the entire tower
- Re-establish links between the horizontal neighbours of each removed position



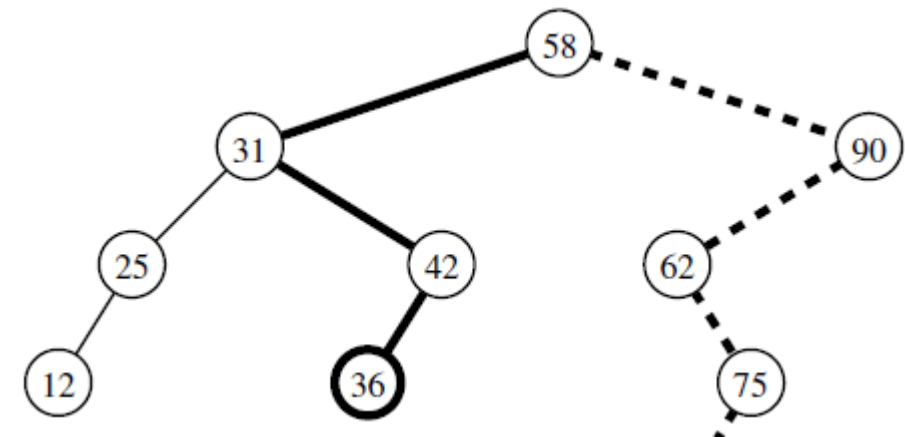


THIS LECTURE

- Search Trees
 - Binary Search Trees
 - Balanced Search Trees

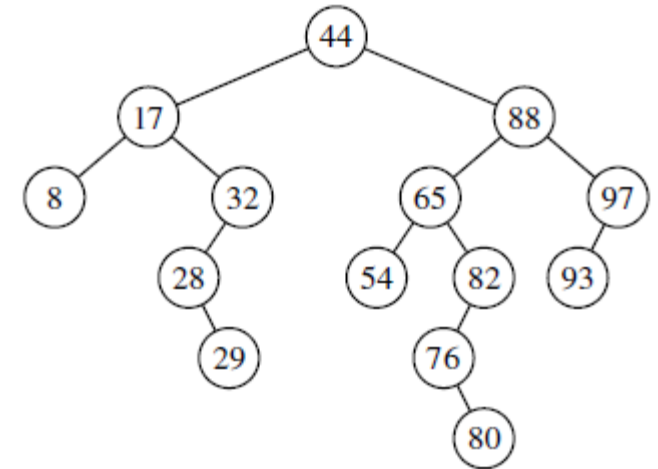
PREVIOUSLY ON BINARY SEARCH TREES

- Binary search tree: using tree to store an ordered sequence of elements in a binary tree
- Let S be a set whose unique elements have an order relation
- A binary search tree for S is a binary tree T such that
 - Position p stores an element of S , denoted as $e(p)$
 - Elements stored in the left subtree of p are less than $e(p)$
 - Elements stored in the right subtree of p are greater than $e(p)$
- Running time for finding an occurrence?
 - Proportional to the height of T
 - How you organize the tree really matters
 - Best case?
 - Worst case?



BINARY SEARCH TREES

- We'd use BSTs to implement a sorted map
 - $M[k]$
 - $M[k] = v$
 - $\text{del } M[k]$
- (Updated) definitions for BST
- Binary Search Tree property: a binary tree T with each position p storing a key-value pair (k, v) such that
 - Keys stored in the left subtree of p are less than k
 - Keys stored in the right subtree of p are greater than k

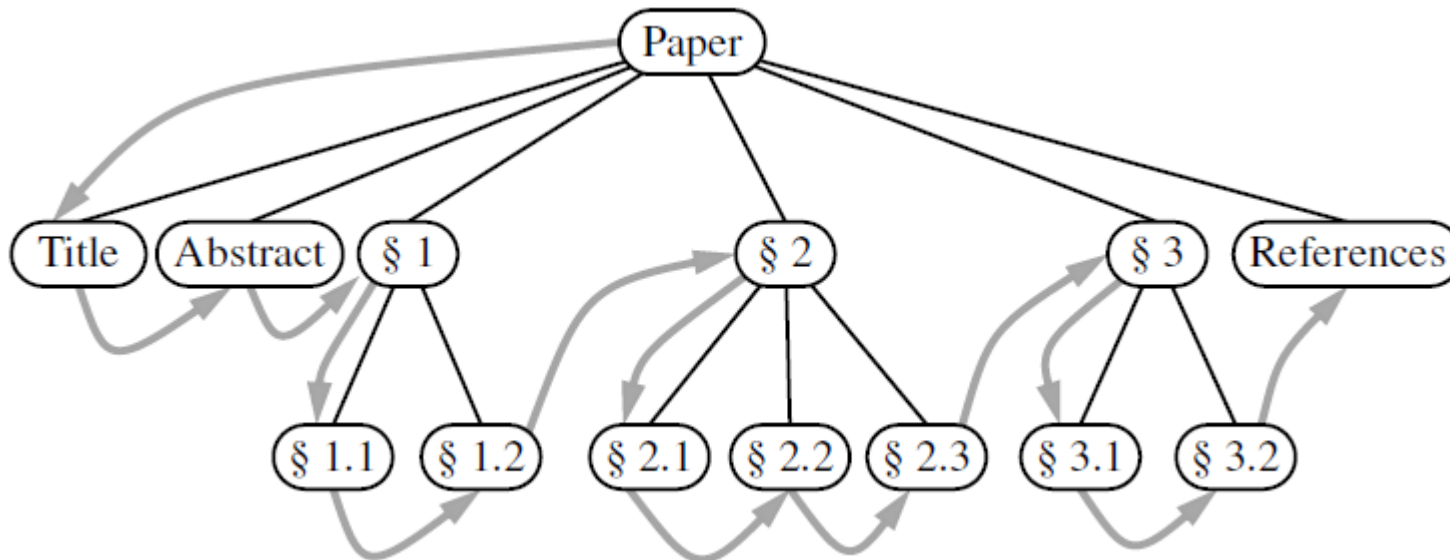


NAVIGATING A BINARY SEARCH TREE

- In order traversal of a binary search tree visits positions in increasing order of their keys
- Proof by induction
 - Base: tree has one item
 - Inductive: recursive inorder traversal: left child(ren) \rightarrow node \rightarrow right child(ren), by binary search tree property, inorder traversal visits positions in increasing order
- Inorder traversal: $O(n) \Rightarrow$ sorted iteration of the keys of a map in $O(n)$, provided that the map is represented as a binary search tree

TREE TRAVERSAL (遍历) ALGORITHMS

- Traversal: a systematic way of accessing or 'visiting' all the positions of T
- Preorder traversal (正序遍历):
 - visit root of T first and then visit the sub-trees recursively
 - If tree is ordered, then the subtrees are traversed according to the order of the children



Algorithm preorder(T, p):
perform the “visit” action for position p
for each child c in $T.children(p)$ **do**
 preorder(T, c)
 {recursion}

TREE ADT

- `T.children(p)`: returns a collection of the children of position `p`
- `T.is_leaf(p)`: returns true if position `p` does not have any children (i.e. it is a leaf node)
- `len(T)`: returns the number of elements in `T`
- `T.is_empty()`: returns true if `T` does not contain any elements
- `T.positions()`: returns a collection of all positions of tree `T`
- `Iter(T)`: returns a iterator of all elements stored in `T`

BINARY TREE ADT

- $T.\text{left}(p)$: returns the position that represents the left child of p
- $T.\text{right}(p)$: returns the position of the right child of p
- $T.\text{sibling}(p)$: returns the sibling of p
- For Linked Structure
 - $T.\text{add_root}(e)$: create a root for an empty tree, store e as the element, and return the position of the root.
 - $T.\text{add_left}(p, e)$: create a node storing element e , link the node as the left child of position p , and return the position
 - $T.\text{replace}(p, e)$: replace the element at position p with element e , and return the previous value stored at p
 - $T.\text{delete}(p)$: remove the node at position p , replace it with its child, if any, and return the element stored previously at p
 - $T.\text{attach}(p, T1, T2)$: attach the internal structure of trees $T1$ and $T2$, as left and right subtrees of leaf position p ; reset $T1$ and $T2$ to empty trees

BINARY SEARCH TREE ADT

- `first()`: returns the position containing the least key, or `None` if the tree is empty
- `last()`: returns the position containing the greatest key, or `None` if empty tree
- `before(p)`: returns the position containing the greatest key that is less than that of position `p`, or `None` if `p` is the first position
- `after(p)`: returns position containing the least key that is greater than that of the position `p`, or `None` if `p` is the last position

BINARY SEARCH TREE ADT

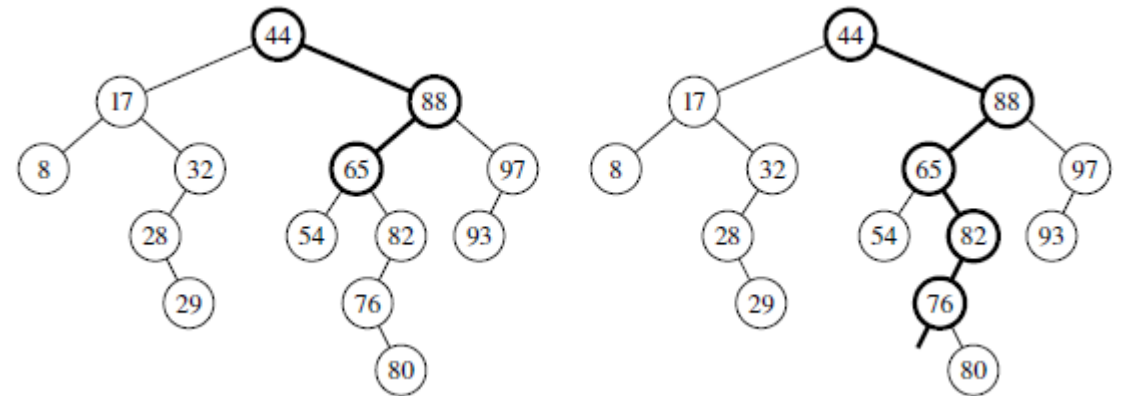
- after() and before() algorithms
- Complexity?
 - Bounded by the height h of the full tree
 - $O(h)$
 - $O(1)$ amortised $\rightarrow n$ calls to after(p) at the first position executes in a total of $O(n)$ time

Algorithm after(p):

```
if right(p) is not None then {successor is leftmost position in p's right subtree}
    walk = right(p)
    while left(walk) is not None do
        walk = left(walk)
    return walk
else {successor is nearest ancestor having p in its left subtree}
    walk = p
    ancestor = parent(walk)
    while ancestor is not None and walk == right(ancestor) do
        walk = ancestor
        ancestor = parent(walk)
    return ancestor
```

SEARCHES

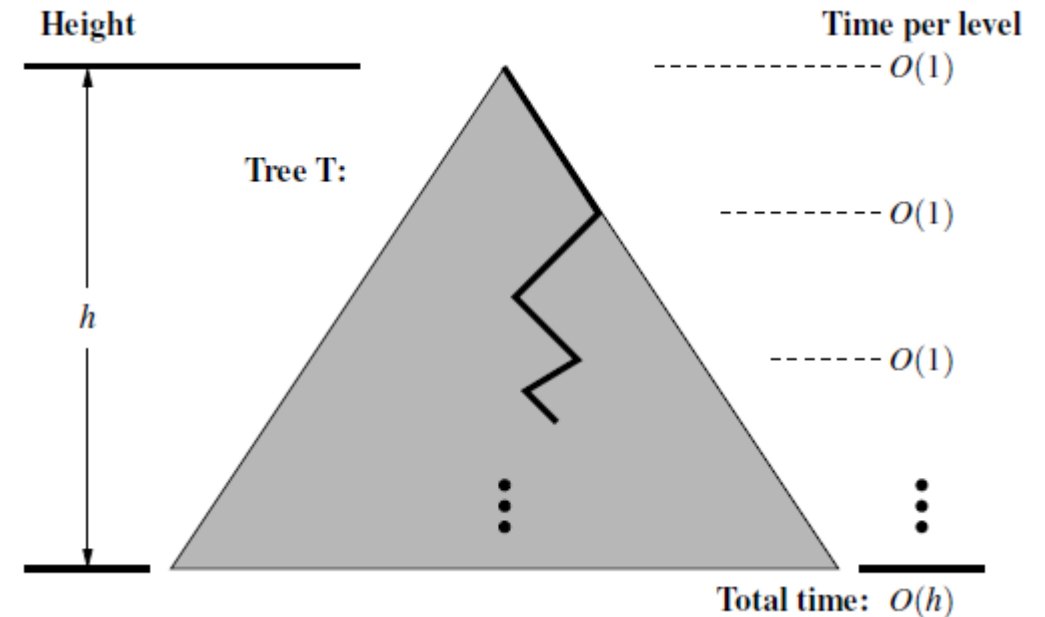
- Locate a particular key by viewing it as a decision tree
- At each position p : is the desired k less than, equal to, or greater than the key stored at position p ?



```
Algorithm TreeSearch( $T, p, k$ ):  
    if  $k == p.key()$  then  
        return  $p$   
    else if  $k < p.key()$  and  $T.left(p)$  is not None then  
        return TreeSearch( $T, T.left(p), k$ )  
    else if  $k > p.key()$  and  $T.right(p)$  is not None then  
        return TreeSearch( $T, T.right(p), k$ )  
    return  $p$ 
```

TREE SEARCHING: ANALYSIS

- TreeSearch: recursive
- Each recursive call of TreeSearch(): made on a child of the previous position
- A Path p : starts at the root and goes down one level at a time.
- Length of path p : bounded by $h+1$, h is the height of T
- TreeSearch: $O(h)$



```
Algorithm TreeSearch( $T, p, k$ ):  
    if  $k == p.key()$  then  
        return  $p$   
    else if  $k < p.key()$  and  $T.left(p)$  is not None then  
        return TreeSearch( $T, T.left(p), k$ )  
    else if  $k > p.key()$  and  $T.right(p)$  is not None then  
        return TreeSearch( $T, T.right(p), k$ )  
    return  $p$ 
```

INSERTIONS

- Map backed by a binary search tree
- $M[k] = v$
 - Search for key k
 - If found, update value
 - Otherwise, create a new node and insert it into the binary search tree
- E.g. insert 68 into tree

Algorithm TreeInsert(T, k, v):

Input: A search key k to be associated with value v

$p = \text{TreeSearch}(T, T.\text{root}(), k)$

if $k == p.\text{key}()$ **then**

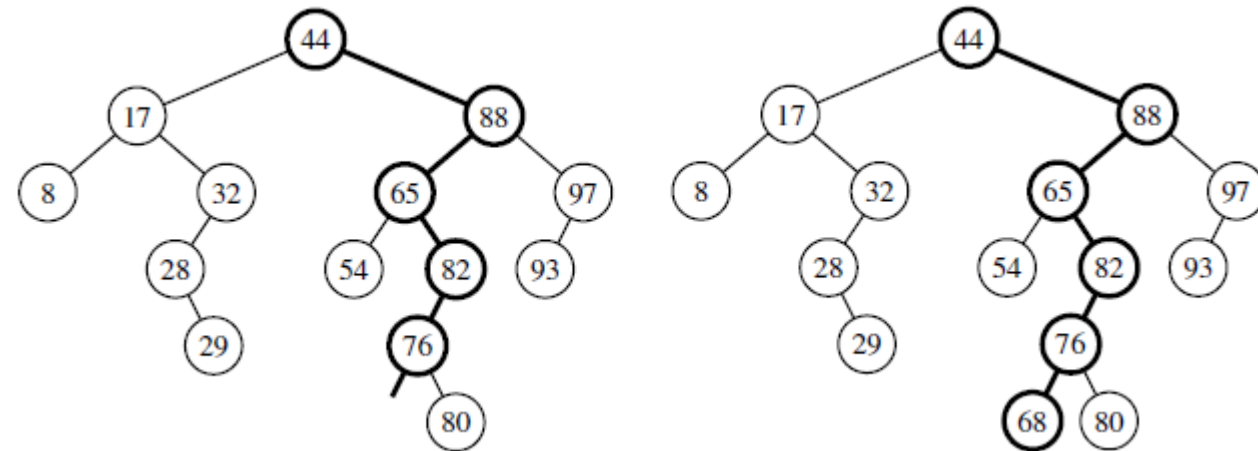
Set p 's value to v

else if $k < p.\text{key}()$ **then**

add node with item (k, v) as left child of p

else

add node with item (k, v) as right child of p

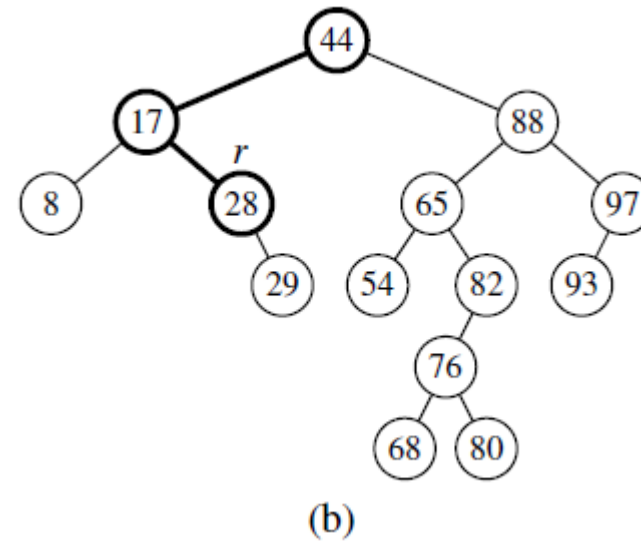
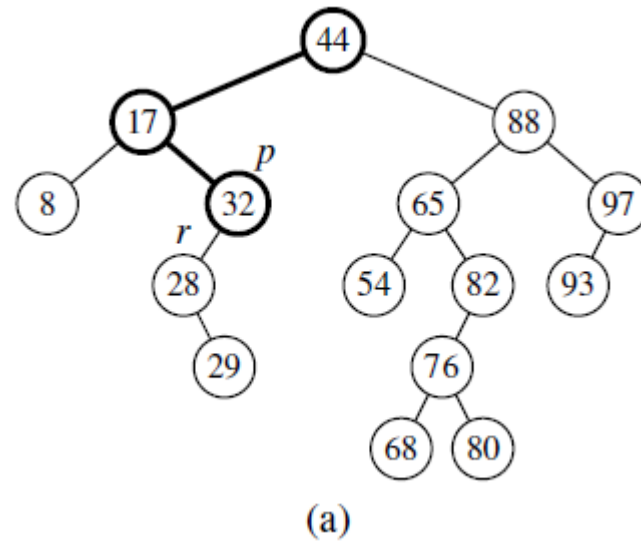


DELETION

- Find the position p of T storing an item with key equal to k , if the search is successful:
 1. If p has at most one child, then delete p , replace it with the child
 2. If p has two children
 - Locate position r having the greatest key that is less than that of p , $r = \text{before}(p)$.
 r is the rightmost position of the left subtree of p
 - Use r 's item as a replacement for position p
 - Delete node at r , since r has at most 1 child, repeat step 1 for r

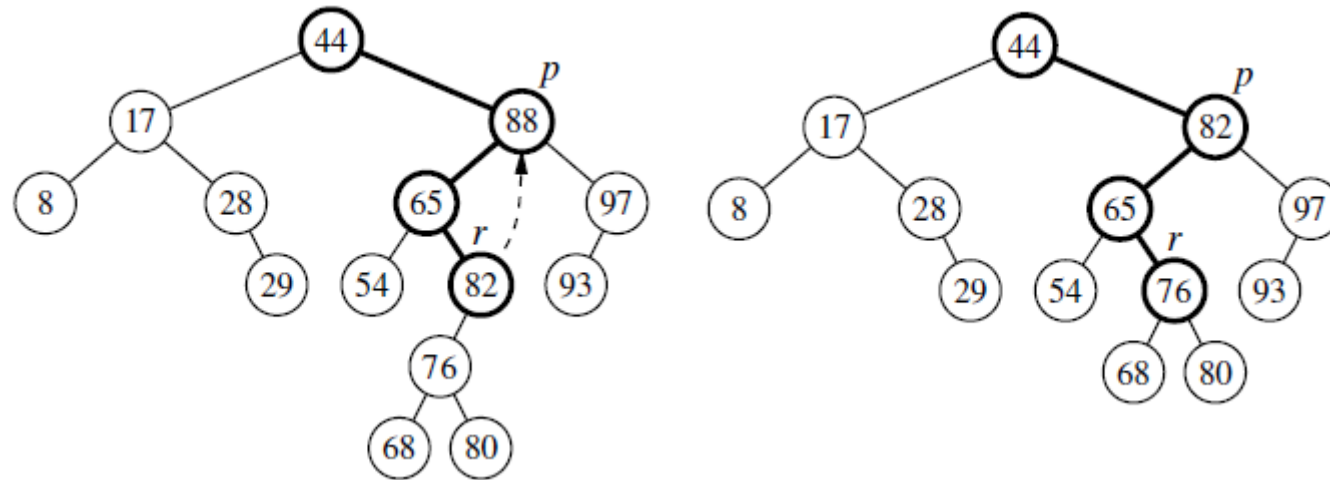
DELETION

- Delete item with $k=32$ with one child r



DELETION

- Delete item with $k=88$



PYTHON IMPLEMENTATION

- Will look at it in a practical

```
1 class TreeMap(LinkedBinaryTree, MapBase):
2     """Sorted map implementation using a binary search tree."""
3
4     #----- override Position class -----
5     class Position(LinkedBinaryTree.Position):
6         def key(self):
7             """Return key of map's key-value pair."""
8             return self.element()._key
9
10        def value(self):
11            """Return value of map's key-value pair."""
12            return self.element()._value
13
14
15    def _subtree_search(self, p, k):
16        """Return Position of p's subtree having key k, or last node searched."""
17        if k == p.key():
18            return p
19        elif k < p.key():
20            if self.left(p) is not None:
21                return self._subtree_search(self.left(p), k)
22            else:
23                if self.right(p) is not None:
24                    return self._subtree_search(self.right(p), k)
25        return p
26
27    def _subtree_first_position(self, p):
28        """Return Position of first item in subtree rooted at p."""
29        walk = p
30        while self.left(walk) is not None:
31            walk = self.left(walk)
32        return walk
33
34    def _subtree_last_position(self, p):
35        """Return Position of last item in subtree rooted at p."""
36        walk = p
37        while self.right(walk) is not None:
38            walk = self.right(walk)
39        return walk
```

PYTHON IMPLEMENTATION

```
40 def first(self):
41     """ Return the first Position in the tree (or None if empty). """
42     return self._subtree_first_position(self.root()) if len(self) > 0 else None
43
44 def last(self):
45     """ Return the last Position in the tree (or None if empty). """
46     return self._subtree_last_position(self.root()) if len(self) > 0 else None
47
48 def before(self, p):
49     """ Return the Position just before p in the natural order.
50
51     Return None if p is the first position.
52     """
53     self._validate(p) # inherited from LinkedBinaryTree
54     if self.left(p):
55         return self._subtree_last_position(self.left(p))
56     else:
57         # walk upward
58         walk = p
59         above = self.parent(walk)
60         while above is not None and walk == self.left(above):
61             walk = above
62             above = self.parent(walk)
63         return above
```

```
65 def after(self, p):
66     """ Return the Position just after p in the
67
68     Return None if p is the last position.
69     """
70     # symmetric to before(p)
71
72 def find_position(self, k):
73     """ Return position with key k, or else neig
74     if self.is_empty():
75         return None
76     else:
77         p = self._subtree_search(self.root(), k)
78         self._rebalance_access(p) # h
79         return p
```

PYTHON IMPLEMENTATION

```
80 def find_min(self):
81     """ Return (key,value) pair with minimum key (or None if empty)
82     if self.is_empty():
83         return None
84     else:
85         p = self.first()
86         return (p.key(), p.value())
87
88 def find_ge(self, k):
89     """ Return (key,value) pair with least key greater than or equal to k
90
91     Return None if there does not exist such a key.
92     """
93     if self.is_empty():
94         return None
95     else:
96         p = self.find_position(k)           # may not find
97         if p.key() < k:                     # p's key is too small
98             p = self.after(p)
99         return (p.key(), p.value()) if p is not None else None

101 def find_range(self, start, stop):
102     """ Iterate all (key,value) pairs such that start <= key < stop.
103
104     If start is None, iteration begins with minimum key of map.
105     If stop is None, iteration continues through the maximum key of map.
106     """
107     if not self.is_empty():
108         if start is None:
109             p = self.first()
110         else:
111             # we initialize p with logic similar to find_ge
112             p = self.find_position(start)
113             if p.key() < start:
114                 p = self.after(p)
115         while p is not None and (stop is None or p.key() < stop):
116             yield (p.key(), p.value())
117             p = self.after(p)
```


PYTHON IMPLEMENTATION

```

118 def __getitem__(self, k):
119     """Return value associated with key k (raise Key
120     if self.is_empty():
121         raise KeyError('Key Error: ' + repr(k))
122     else:
123         p = self._subtree_search(self.root(), k)
124         self._rebalance_access(p) # hook for
125         if k != p.key():
126             raise KeyError('Key Error: ' + repr(k))
127         return p.value()
128
129 def __setitem__(self, k, v):
130     """Assign value v to key k, overwriting existing
131     if self.is_empty():
132         leaf = self._add_root(self._Item(k,v))
133     else:
134         p = self._subtree_search(self.root(), k)
135         if p.key() == k:
136             p.element()._value = v # replace
137             self._rebalance_access(p) # hook for
138             return
139         else:
140             item = self._Item(k,v)
141             if p.key() < k:
142                 leaf = self._add_right(p, item) # inherit
143             else:
144                 leaf = self._add_left(p, item) # inherit
145             self._rebalance_insert(leaf) # hook for

```

```

147 def __iter__(self):
148     """Generate an iteration
149     p = self.first()
150     while p is not None:
151         yield p.key()
152         p = self.after(p)
153
154 def delete(self, p):
155     """Remove the item at given Position."""
156     self._validate(p) # inhe
157     if self.left(p) and self.right(p): # p has
158         replacement = self._subtree_last_position(s
159         self._replace(p, replacement.element())
160         p = replacement
161         # now p has at most one child
162         parent = self.parent(p)
163         self._delete(p) # inhe
164         self._rebalance_delete(parent) # if ro
165
166 def __delitem__(self, k):
167     """Remove item associated with key k (raise
168     if not self.is_empty():
169         p = self._subtree_search(self.root(), k)
170         if k == p.key():
171             self.delete(p) # rely
172             return # succ
173             self._rebalance_access(p) # hook
174         raise KeyError('Key Error: ' + repr(k))

```

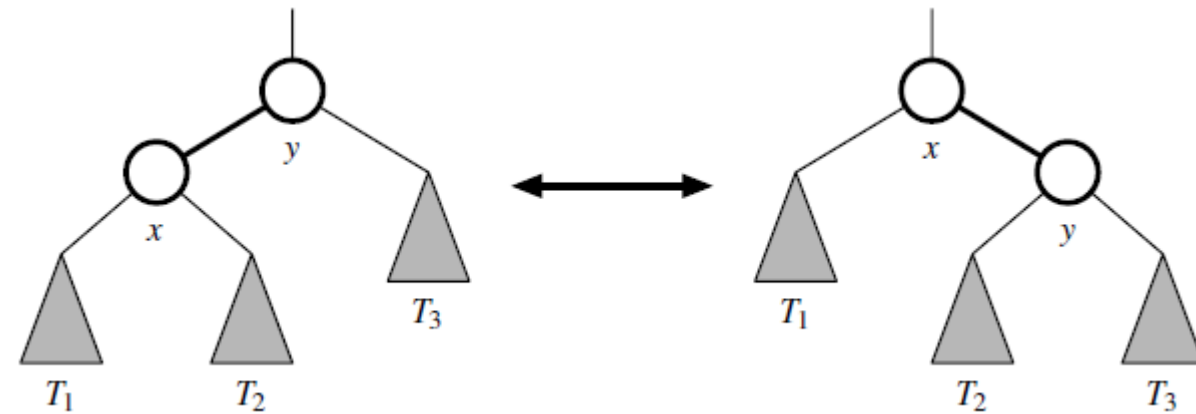
PERFORMANCE OF A BINARY SEARCH TREE

- Almost all operations have a worst-case running time of $O(h)$
- Single call to `after()` is worst case $O(h)$, n successive calls made during a call to `__iter__` require a total of $O(n)$ time" each edge is traced at most twice
- $O(1)$ amortised time bounds
- Is $O(h)$ same as $O(\log n)$?
- No, BST can be unbalanced

Operation	Running Time
$k \text{ in } T$	$O(h)$
$T[k], T[k] = v$	$O(h)$
$T.\text{delete}(p), \text{del } T[k]$	$O(h)$
$T.\text{find_position}(k)$	$O(h)$
$T.\text{first}(), T.\text{last}(), T.\text{find_min}(), T.\text{find_max}()$	$O(h)$
$T.\text{before}(p), T.\text{after}(p)$	$O(h)$
$T.\text{find_lt}(k), T.\text{find_le}(k), T.\text{find_gt}(k), T.\text{find_ge}(k)$	$O(h)$
$T.\text{find_range}(\text{start}, \text{stop})$	$O(s + h)$
$\text{iter}(T), \text{reversed}(T)$	$O(n)$

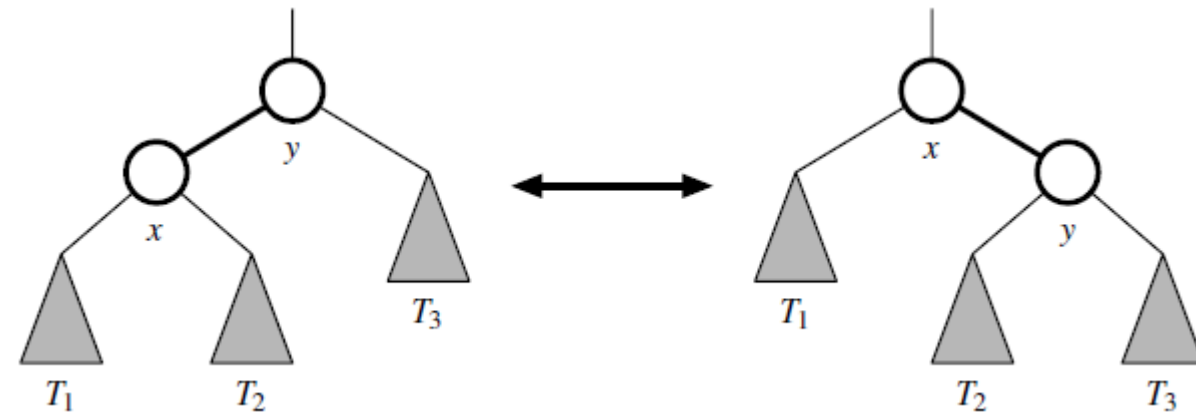
BALANCED SEARCH TREES

- Balanced binary search tree: $O(\log n)$ time for basic map operations
- What about the running time of operations after some sequence of operations?
 - $O(n)$
 - Why?
- Balanced Search Trees: stronger performance guarantees
- Main idea: rotation



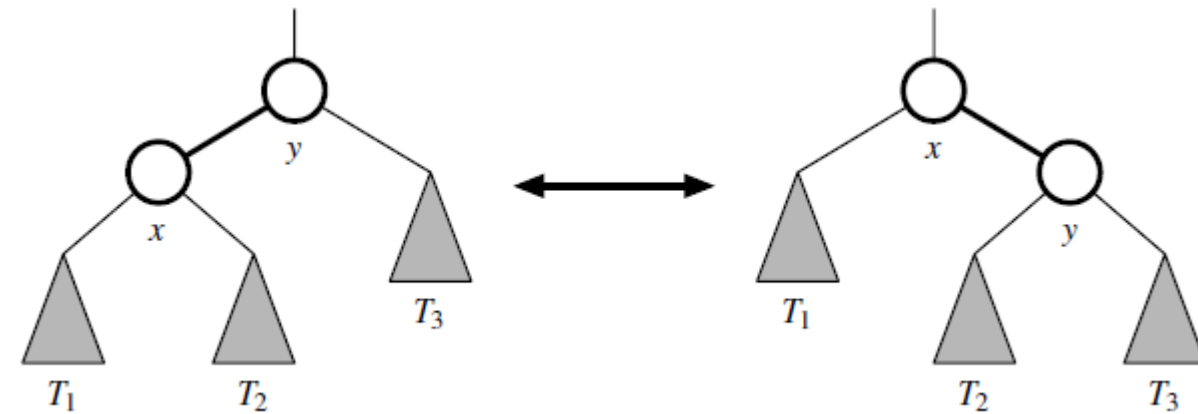
BALANCED SEARCH TREES

- Before Rotation
 - Position x : $x.\text{key} < y.\text{key}$
 - For each node in T_1 , their keys are less than $x.\text{key}$
 - For each node in T_2 , their keys are greater than $x.\text{key}$, but less than $y.\text{key}$
- After rotation
 - Position x : $x.\text{key} < y.\text{key}$
 - For each node in T_1 , their keys are less than $x.\text{key}$
 - For each node in T_2 , their keys are greater than $x.\text{key}$, but less than $y.\text{key}$



BALANCED SEARCH TREES

- Single rotation: a constant number of parent-child relationships are modified
 - $O(1)$ for linked binary with a linked binary tree representation
- Rotations allow the shape of a tree to be modified while maintaining the search tree property
 - Rightward rotation: depth of each node in T_1 reduced by 1, depth of each node in T_3 increased by 1
- One or more rotation: trinode restructuring



BALANCED SEARCH TREES

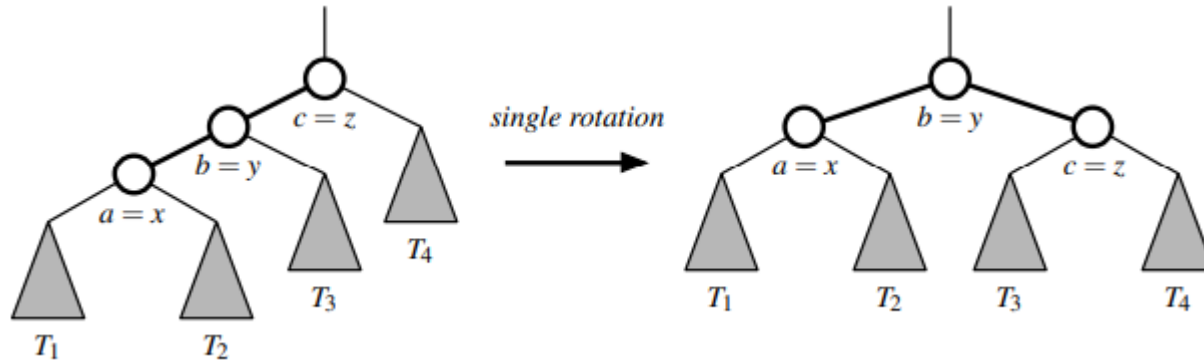
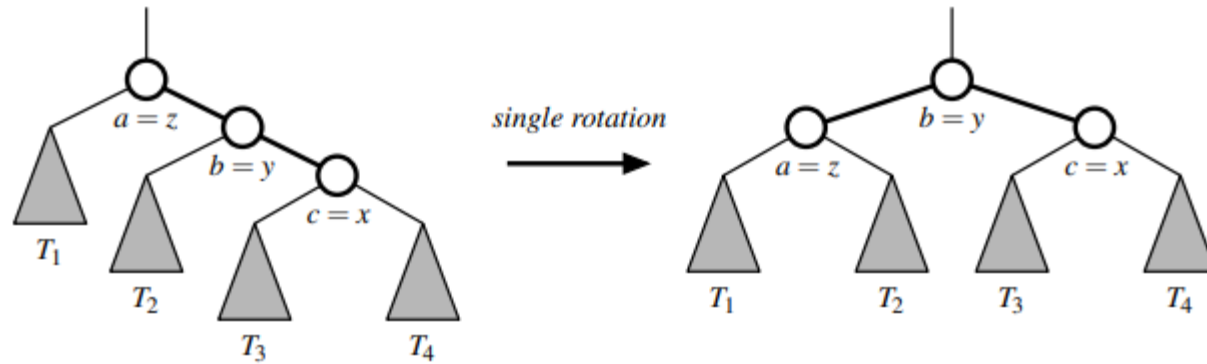
Algorithm restructure(x):

Input: A position x of a binary search tree T that has both a parent y and a grandparent z

Output: Tree T after a trinode restructuring (which corresponds to a single or double rotation) involving positions x , y , and z

- 1: Let (a, b, c) be a left-to-right (inorder) listing of the positions x , y , and z , and let (T_1, T_2, T_3, T_4) be a left-to-right (inorder) listing of the four subtrees of x , y , and z not rooted at x , y , or z .
- 2: Replace the subtree rooted at z with a new subtree rooted at b .
- 3: Let a be the left child of b and let T_1 and T_2 be the left and right subtrees of a , respectively.
- 4: Let c be the right child of b and let T_3 and T_4 be the left and right subtrees of c , respectively.

BALANCED SEARCH TREES



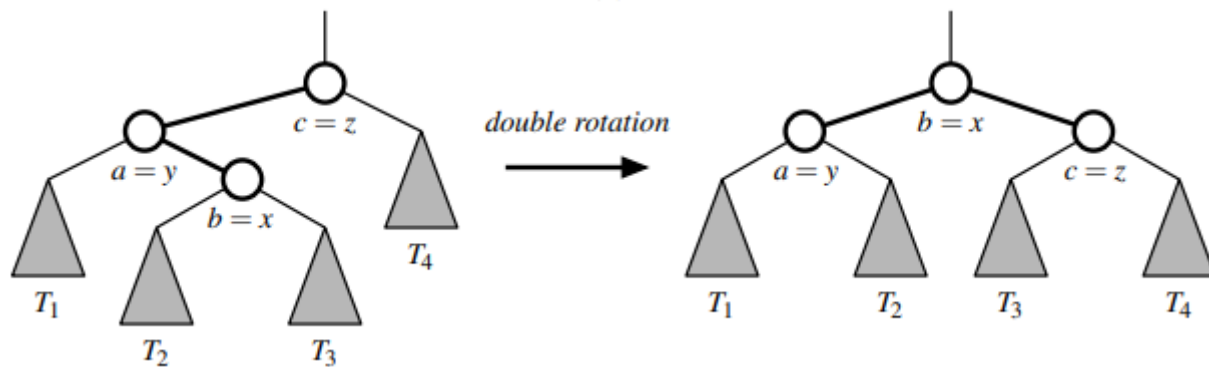
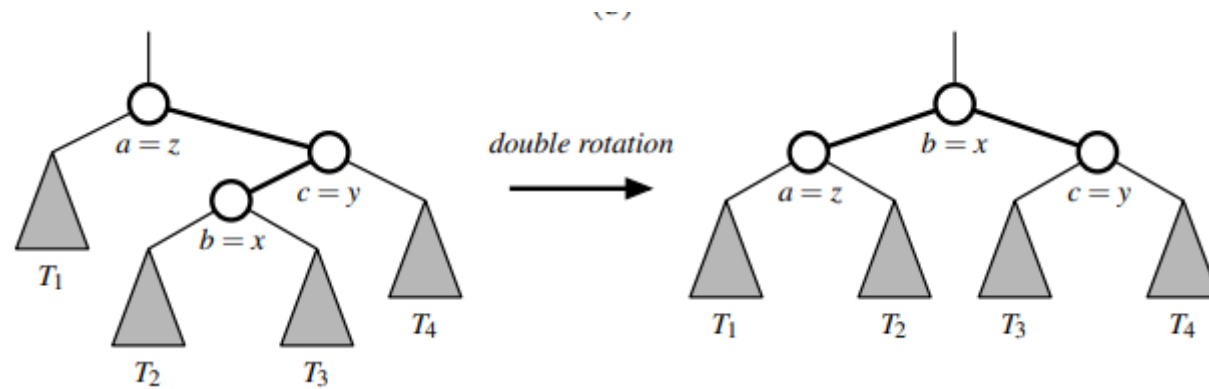
Algorithm restructure(x):

Input: A position x of a binary search tree T that has both a parent y and a grandparent z

Output: Tree T after a trinode restructuring (which corresponds to a single or double rotation) involving positions x , y , and z

- 1: Let (a, b, c) be a left-to-right (inorder) listing of the positions x , y , and z , and let (T_1, T_2, T_3, T_4) be a left-to-right (inorder) listing of the four subtrees of x , y , and z not rooted at x , y , or z .
- 2: Replace the subtree rooted at z with a new subtree rooted at b .
- 3: Let a be the left child of b and let T_1 and T_2 be the left and right subtrees of a , respectively.
- 4: Let c be the right child of b and let T_3 and T_4 be the left and right subtrees of c , respectively.

BALANCED SEARCH TREES



Algorithm restructure(x):

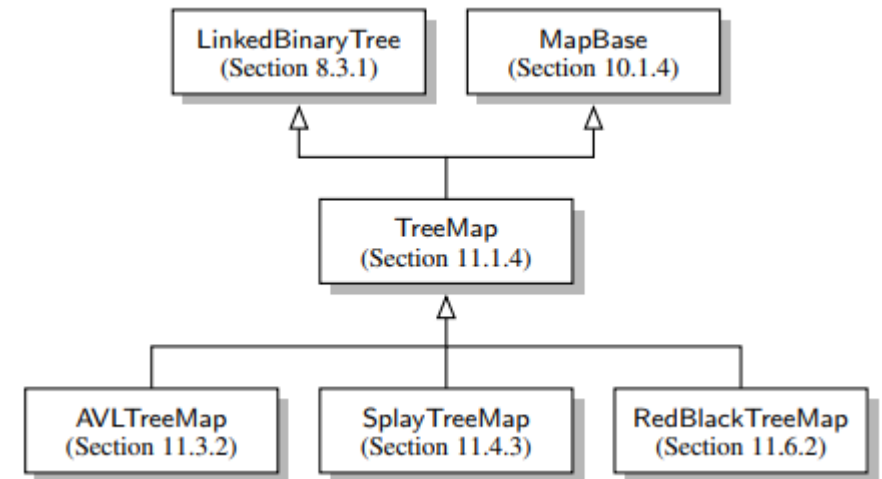
Input: A position x of a binary search tree T that has both a parent y and a grandparent z

Output: Tree T after a trinode restructuring (which corresponds to a single or double rotation) involving positions x , y , and z

- 1: Let (a, b, c) be a left-to-right (inorder) listing of the positions x , y , and z , and let (T_1, T_2, T_3, T_4) be a left-to-right (inorder) listing of the four subtrees of x , y , and z not rooted at x , y , or z .
- 2: Replace the subtree rooted at z with a new subtree rooted at b .
- 3: Let a be the left child of b and let T_1 and T_2 be the left and right subtrees of a , respectively.
- 4: Let c be the right child of b and let T_3 and T_4 be the left and right subtrees of c , respectively.

PYTHON FRAMEWORK FOR BALANCING SEARCH TREES

- `_rebalance_insert(p)`: called by `__setitem__`
- `_rebalance_delete(p)`: called by `__delitem__`
- `_rebalance_access(p)`: called by `__getitem__`



PYTHON FRAMEWORK FOR BALANCING SEARCH TREES

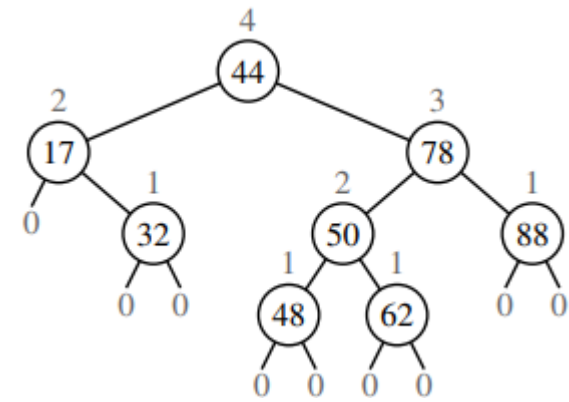
```
177 def _relink(self, parent, child, make_left_child): 186
178     """Relink parent node with child node (we all 187
179     if make_left_child: # make 188
180         parent._left = child 189
181     else: # make 190
182         parent._right = child 191
183     if child is not None: # make 192
184         child._parent = parent 193
```

```
194
195     self._relink(z, x, y == z._left) 204
196     # now rotate x and y, including t 205
197     if x == y._left: 206
198         self._relink(y, x._right, True) 207
199         self._relink(x, y, False) 208
200     else: 209
201         self._relink(y, x._left, False) 210
202         self._relink(x, y, True) 211
```

```
212
213     self._relink(z, x, y == z._left) 213
214     # now rotate x and y, including t 214
215     if x == y._left: 214
216         self._relink(y, x._right, True) 215
217         self._relink(x, y, False) 216
218     else: 217
219         self._relink(y, x._left, False) 218
220         self._relink(x, y, True) 219
221     return x 220
222
223 def _restructure(self, x): 221
224     """Perform trinode restructure of Position x with 222
225     y = self.parent(x) 223
226     z = self.parent(y) 224
227     if (x == self.right(y)) == (y == self.right(z)): 225
228         self._rotate(y) 226
229         return y 227
230     else: 228
231         self._rotate(x) 229
232         self._rotate(x) 230
233         return x 231
```

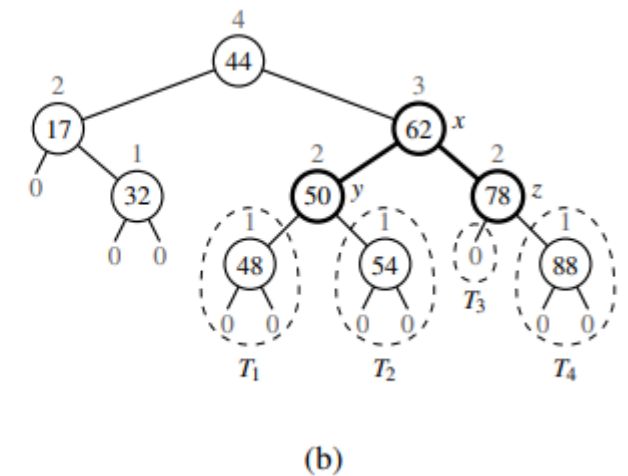
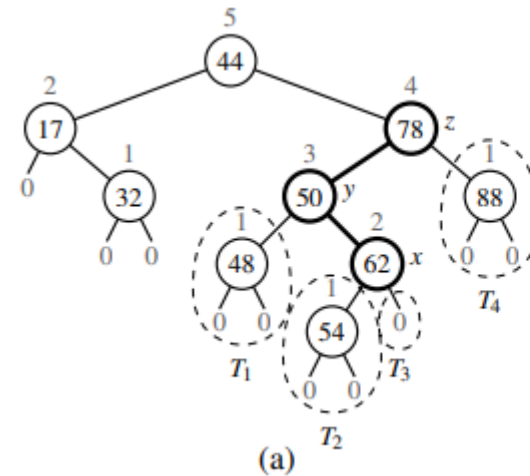
AVL TREES

- AVL: Adelson-Velsky and Landis
- Adds a rule to the binary search tree to maintain a logarithmic height for the tree
- Height: number of edges on the longest path vs number of nodes on this longest path
 - Leaf position has height 1
- **Height balance property:** for every position p of T , the heights of the children of p differ by at most 1
- A subtree of an AVL tree is itself an AVL tree
- The height of an AVL tree storing n entries is $O(\log n)$

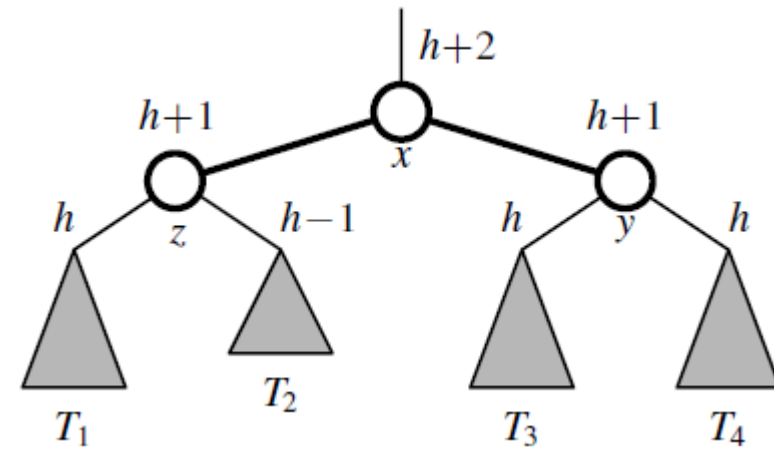
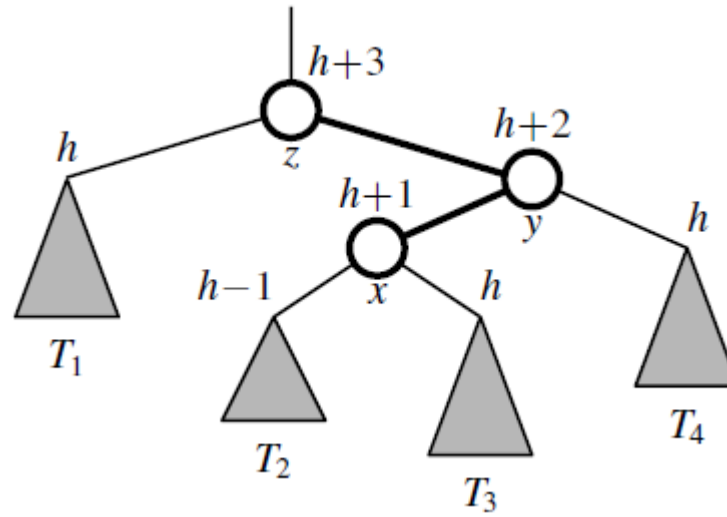
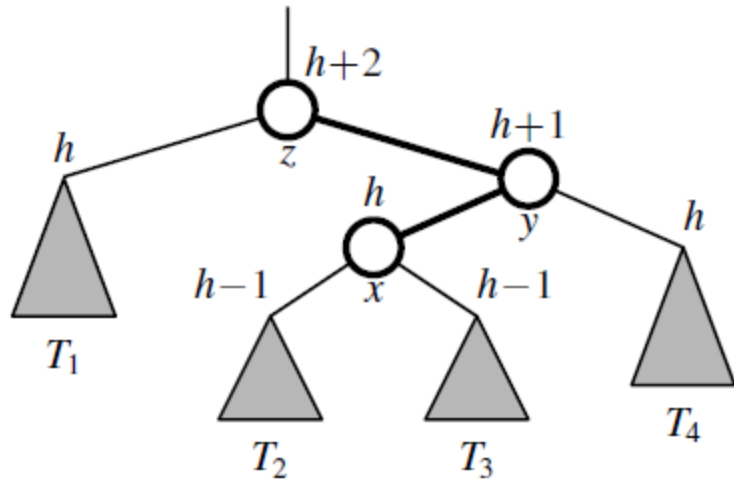


AVL TREES

- Insertion: insert item with key 54
- “search and repair”: going up from p to the root of T
 - Z: first unbalanced position
 - Y: child of z with higher height, y must be an ancestor of p
 - X: child of y with higher height (no tie, x must be an ancestor of p)
 - Call the trinode restructuring method, restructure(x)

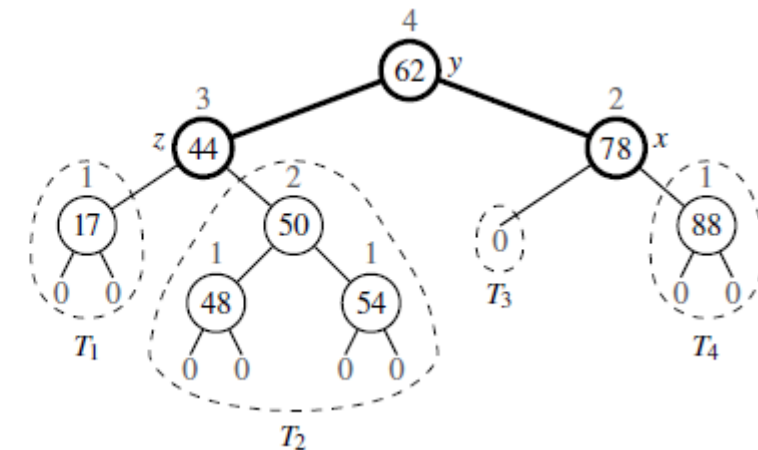
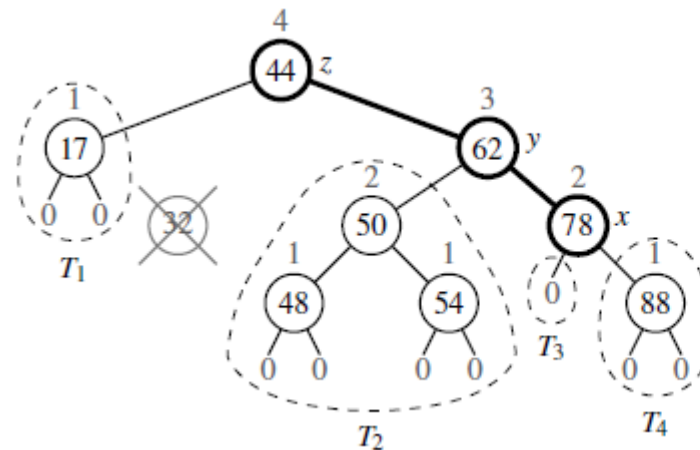


AVL TREES



AVL TREES

- Deletion: delete item with key 32
- Trinode restructuring:
 - Z: first unbalanced position
 - Y: child of z with larger height
 - X: child of y, such that if one the children of y is taller than the other, let x be the taller child of y, else let x be the child of y on the same side as y
 - Perform restructure(x)



PERFORMANCE OF AVL TREES

Operation	Running Time
$k \text{ in } T$	$O(\log n)$
$T[k] = v$	$O(\log n)$
$T.\text{delete}(p)$, $\text{del } T[k]$	$O(\log n)$
$T.\text{find_position}(k)$	$O(\log n)$
$T.\text{first}()$, $T.\text{last}()$, $T.\text{find_min}()$, $T.\text{find_max}()$	$O(\log n)$
$T.\text{before}(p)$, $T.\text{after}(p)$	$O(\log n)$
$T.\text{find_lt}(k)$, $T.\text{find_le}(k)$, $T.\text{find_gt}(k)$, $T.\text{find_ge}(k)$	$O(\log n)$
$T.\text{find_range}(\text{start}, \text{stop})$	$O(s + \log n)$
$\text{iter}(T)$, $\text{reversed}(T)$	$O(n)$



THANKS

See you in the next session!